SOME CONTRIBUTIONS TO THE CHANGE POINT PROBLEM

CHITHRAN VADAVERKKOT VASUDEVAN

Some Contributions to the Change Point Problem

by

© Chithran Vadaverkkot Vasudevan

A thesis submitted to the

School of Graduate Studies

in partial fulfillment of the

requirements for the degree of

Master of Science in Statistics

Department of Mathematics and Statistics

Memorial University of Newfoundland

05 May 2011

St. John's Newfoundland

Abstract

The identification of changes in process parameters is an important statistical problem in industrial-process monitoring. The existing methods, the change point model (Hawkins et al. (2003)) and the modified information criterion (Chen et al. (2006)) rely on the parametric distribution of the quality characteristic, and any deviation from the specified model may lead to incorrect conclusions. We propose an empirical-likelihood-based information criterion (ELIC) for identifying changes in the process parameters. The main advantage of our method is that we do not need to specify a parametric distribution for the quality characteristic. Our simulation studies indicate that our method is as good as existing methods when the distribution of the quality characteristic is known, and it outperforms existing methods when the distribution is approximated or misspecified. We introduce the EM test in the Bayesian approach for the change point problem suggested by Bansal et al. (2008). From simulation studies, we see that the Bayesian EM test performs as well as the Bayesian approach with full EM iteration. We compare the performance of all methods for identifying the change point in a wide range of data scenarios. Our methods are applied to two case studies.

Acknowledgements

By the blessings of God, it is my privilege to express my deepest gratitude to my supervisor Dr. Asokan Mulayath Variyath for his immense support in every way for my Masters program. I am thankful to him for introducing me to the wonderful world of statistical process control and empirical likelihood. With his encouragement, endless patience, and constant support, I was able to improve my research skills. I was blessed to have him as my supervisor and am grateful for his guidance in the preparation of my thesis.

I want to sincerely acknowledge the financial support provided by the School of Graduate Studies, the Department of Mathematics and Statistics, and Dr. A.M. Variyath in the form of a graduate assistantship and teaching assistantships. Further, I express my sincere thanks to the lovely and helpful staff at our departmental office for providing necessary facilities.

I am grateful to Dr. P.G. Sankaran, Dr. N. Balakrishna (Cochin University of Science and Technology, India), and Prof. P.K. Venugopalan (Sree Kerala Varma College, Thrissur, India) for referring me to the Memorial University of Newfoundland. I take this opportunity to thank Dr. Zhao Zhi Fan, Dr. Brajendra Sutradhar, Dr.

Asokan Mulayath Variyath, and Dr. J.C. Loredo-Osti for introducing me to various fields in statistics through their courses.

I thank all the graduate students in mathematics and statistics for helping me to complete this program. The calm and quiet atmosphere of St. John's helped me to perform well throughout my program. I am grateful for the support of my friends here at St. John's and also back home.

Last but not least I would like to thank my parents and other family members for their unconditional love and support.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Quality has become an important consumer decision factor. Even long-term customers may switch to another manufacturer if it can deliver better quality. The increased customer demand for high-quality products and services forces manufacturers to improve quality on a continuous basis. Modern quality-management methods stress the need for the application of statistical techniques to improve quality and productivity.

The introduction of ISO 9000 quality systems as well as the six-sigma methodology calls for the application of statistical methods to monitor and control the process. Statistical process control (SPC) is a set of pragmatic tools widely used in industry to identify special causes and to bring the process under the influence of a stable

system of chance causes. The basic philosophy of SPC is that a product coming out of a controlled process will be of good quality. The timely identification of changes in the process is necessary so that appropriate corrective and preventive actions can be taken to improve quality and productivity.

The *Change Point Problem* involves the identification of changes at unknown times and locating these changes in stochastic processes. In industrial-process monitoring, the identification of the change point for the process mean is of great research interest. Traditionally, control charts are used to monitor industrial processes. Control charts, developed by Dr. W.A. Shewart in the 1920s (see Shewart (1931) and Shewart (1939)) while working for Bell Laboratories, are a technically sophisticated SPC tool. A control chart is a graphical tool that detects changes in the manufacturing conditions by comparing the observed values with limits derived from past experience. Control charts are used to check whether or not the process is statistically under control. In practice, control charts are implemented in two stages. Consecutive samples of a fixed size are collected at regular intervals and the quality characteristics of interest are measured. Using these data, also referred to as phase I data, control limits are estimated. These control limits are then used for monitoring the future data, referred to as phase II data, collected via the same procedure. The timely use of control charts helps to detect changes in the quality characteristics due to assignable causes.

The removal of assignable causes will reduce variability and improve the process by bringing it back under the influence of chance causes alone.

Commonly used Shewart control charts are $\bar{X} - R$ or $\bar{X} - S$ charts for measured data and the proportion defective (P) chart for attribute data. Control charts are effective at process monitoring. They efficiently detect isolated abnormal points and quickly detect major changes. They also indicate outliers. One of the major drawbacks of Shewart control charts is that the statistic plotted in the chart is based only on the corresponding subgroup sample measurements, i.e., it does not utilize the information available from the previously observed samples. Thus, Shewart control charts are insensitive to subtle changes in the process (see Ryan (2000)). More specifically, Shewart control charts are less effective when the magnitude of the shift is less than 1.5 times the standard deviation. Cumulative sum (CUSUM) charts and exponentially weighted moving average (EWMA) charts are widely used to detect small changes in the process (see Ryan (2000) and Hawkins and Olwell (1998)).

1.1 CUSUM Chart

The cumulative sum (CUSUM) chart is widely used for detecting small shifts in a process; it was proposed by Page (1954). To implement conventional CUSUM charts,

we need an estimate of the true values of the in-control mean and standard deviation. CUSUM computes a cumulative sum of the measurements. In CUSUM charts, we focus on two kinds of process shifts: an upward shift in the mean ($\mu_2 > \mu_1$) and a downward shift in the mean ($\mu_1 > \mu_2$), where μ_1 is the process mean before the shift and μ_2 is the process mean after the shift in the process. Let $X_1, X_2, \ldots, X_n$ be independently and identically distributed (iid) observations of size n from $N(\mu, \sigma^2)$. For an upward shift, define

$$S_0^+ = 0$$

$$S_n^+ = \max(0, S_{n-1}^+ + X_n - k), \quad \text{n} = 1, 2, \ldots$$

where $k = (\mu_1 + \mu_2)/2$. If $S_n^+ > h$, where h is the critical value and a function of μ_1 and σ^2, then we conclude that an upward shift in the process mean has occurred. For a downward shift, define

$$S_0^- = 0$$

$$S_n^- = \min(0, S_{n-1}^+ + X_n + k), \quad n = 1, 2, \ldots$$

The chart will signal a downward shift in the process mean if $S_n^- < -h$. The main limitation of this approach is that we should know the estimated values of μ_1, μ_2, and σ^2 in advance. Another approach is to construct a self-starting CUSUM chart where we do not require these estimates in advance. Let $\bar{X}_j$ and s_j^2 be the sample mean and sample variance of the first j observations. Compute $t_j = \sqrt{\frac{j-1}{j}} \frac{X_j - \bar{X}_{j-1}}{s_{j-1}}$

and for each t_j, define $U_j = \Phi^{-1}[T_{j-2}(t_j)]$, where T_{j-2} is the cumulative distribution function of the Student's t-distribution with $(j-2)$ degrees of freedom, and Φ is the standard normal cumulative distribution function. The random variable U_j has an exact standard normal distribution for all $j > 2$ and the U_j's are statistically independent. As n increases, $\bar{X}_j \to \mu_1$, $s_n \to \sigma$, and $U_n \to \frac{X_n - \mu_1}{\sigma} \sim N(0,1)$. We now have a known-parameter CUSUM chart of standard normal-distribution quantities. The recursions of the upward and downward CUSUMs for the mean are defined to be

$$S_n^+ = \max(0, S_{n-1}^+ + U_n - k), \quad n = 1, 2, \ldots$$

$$S_n^- = \min(0, S_{n-1}^- + U_n + k), \quad n = 1, 2, \ldots$$

where k is the CUSUM reference value, $S_0^+ = 0$, and $S_0^- = 0$. Let $h_{n,\alpha}$ be the critical value of the CUSUM for a particular sample size n and false-alarm rate α. A shift in the process mean occurs when $S_n^+ > h_{n,\alpha}$ or $S_n^- < -h_{n,\alpha}$. Hawkins and Olwell (1998) have derived various combinations of k and $h_{n,\alpha}$ in relation to the in-control average run length (ARL). The CUSUM charts are not as efficient as Shewart control charts in detecting detect large shifts in the process mean. Also, special patterns in the observed values are hard to detect in CUSUM charts (see the National Institute of Standards and Technology (NIST)/SEMATECH e-Handbook).

1.2 EWMA Chart

The exponentially-weighted moving average (EWMA) chart is also widely used for detecting small shifts in the mean; it was introduced by Roberts (1959). The EWMA chart is almost the same as the CUSUM chart but it is easier to operate.

In an EWMA chart we plot the statistic w_t, which is defined to be

$$w_t = \lambda X_t + (1-\lambda)w_{t-1}, \quad t = 1, 2, \ldots$$

where λ $(0 < \lambda \leq 1)$ is the weight, and the initial value is $w_0 = \bar{X}$. X_t is the observation at time point t. The above expression is a weighted average in which X_t has weight λ and the value of the expression at time $t-1$ has weight $1-\lambda$.

The variance of the weighted average w_t is given by

$$\sigma^2_{w_t} = \sigma^2 \left(\frac{\lambda}{2-\lambda}\right)\left[1-(1-\lambda)^{2t}\right]$$

Here σ^2 is estimated as $\dfrac{\overline{MR}}{d_2}$, where $\overline{MR}$ is the average of the moving ranges of order 2, and d_2 is the constant used to estimate σ. Since we are considering a moving average of order 2, $d_2 = 1.128$. The control limits for w_t are

$$\bar{X} \pm 3\sigma_{w_t}$$

If $\lambda \geq 0.2$, however, $(1-\lambda)^{2t}$ will be close to zero for all $t \geq 5$, so $\sigma^2_{w_t}$ can be

approximated as $\sigma^2(\frac{\lambda}{2-\lambda})$. Then the control limits can be approximated as

$$\bar{X} \pm 3\hat{\sigma}\sqrt{\frac{\lambda}{2-\lambda}}$$

The pattern of points in this control chart gives vital information about the shift in the process mean. EWMA charts are also not effective at identifying large shifts. They are also inefficient for detecting outliers (see NIST/SEMATECH e-Handbook).

None of these methods are efficient at signaling shifts in the process mean. This has led researchers to develop efficient methods for identifying changes in process parameters.

1.3 Change Point Problem

There are few promising methods in the literature for the identification of change points. Methods such as the Hawkins method (the change point model) proposed by Hawkins, Qiu and Kang (2003), the modified information criterion (MIC) by Chen, Gupta and Pan (2006), and the Bayesian approach for the change point problem by Bansal, Du and Hamadani (2008) are commonly used. The Hawkins method assumes that the quality characteristic is normally distributed. It performs a two-sample t test by assuming one data point to be a change point. This procedure is repeated for each data point as a candidate change point. The data point with the most significant test

statistic is identified as the change point. MIC is based on the two-sample likelihood ratio test and a modification of Bayesian information criterion (BIC). Here too the data point with the largest information criterion value is identified as the change point. In the Bayesian approach, the identification of the change point is based on likelihood maximization through the EM algorithm. It provides a weighted likelihood for each data point, and maximizing the sum of these likelihoods leads to one of the weights tending to 1. The corresponding position is the change point. Note that all three methods rely on the parametric distribution of the quality characteristic of interest.

In many situations, the distribution of the observations is not well defined. It is also difficult to assess the exact distribution of quality characteristics such as runout (ovality), strength, etc. Consider the manufacturing of an engine valve for automobiles. The circular runout (ovality) of the head-stock side of the valve is an important quality characteristic. A high runout indicates poor performance, and it is recommended that the runout be monitored on a continuous basis. A histogram of sample runout data is shown in Fig. 5.1 (page 75) and it can be seen that the runout distribution is not normal. Using existing methods with the assumption of a misspecified or approximate distribution may lead to an incorrect adjustment of the process. This limits the application of these methods since in many situations we are unsure about

the distribution of the quality characteristic. Therefore, a method which does not depend on the distribution of the observations is preferable. We propose a nonparametric likelihood-based information criterion for identifying the change point.

The empirical likelihood (EL), proposed by Owen (1988, 2001), has similar properties to its counterpart, parametric likelihood. This data-driven approach has been extended to estimation theory and many other statistical applications such as linear models, survival analysis, and sampling. Confidence intervals based on EL have data-driven shapes. More interestingly, the estimation of scale parameters is not required. Moreover, EL-based methods are more robust to a slight misspecification of the parametric assumptions. Our proposed method is a modification of the method suggested by Chen et al. (2006); we replace the parametric likelihood by the EL based on a set of estimating equations. We use the adjusted EL proposed in Variyath (2006) and Chen et al. (2008) to avoid the nonexistence of solutions when computing the EL ratio statistic. We compute the EL-based information criterion for each data point assuming it to be a change point, and the data point with the maximum EL information criterion is selected as the change point. Our simulation studies show that the EL-based criterion performs as well as MIC if the parametric assumptions are correct. When there is some misspecification of the distributional assumption, our proposed method has a clear advantage over MIC. We modified the Bayesian

approach to change point problem by introducing EM test, proposed by Li, Chen and Marriott (2009) and Chen and Li (2009). We apply our proposed methodologies to real case studies.

In the next chapter, we briefly discuss the existing approaches for the identification of the change point, such as the Hawkins method and the MIC method. In Chapter 3, we briefly review EL, introduce our proposed EL-based information criterion (ELIC) for change point problems, and present an algorithm and a performance analysis. In Chapter 4, we review the Bayesian approach for change point problems and introduce the EM test for the Bayesian approach and assess its performance. In Chapter 5, we apply our methods to two case studies. Our concluding remarks are provided in Chapter 6.

Chapter 2

Review of Change Point Models

In change point analysis, we wish to identify the time at which a process exhibits changes in the process parameters. From a statistical point of view, the change point can be defined as a reference point classifying the sequence of observations into two groups: those before and those after the reference point. These two groups have different distributions because of the changes in the values of one or more parameters. Let $X_1, X_2, \ldots, X_n$ be a sequence of independent random variables with probability densities

$$f(x_i; \theta_1) \text{ for } i = 1, 2, 3, \ldots, \tau$$

$$f(x_i; \theta_2) \text{ for } i = \tau + 1, \ldots, n,$$

where θ_1 and θ_2 are the parameters before and after the unknown change point τ and $f(.)$ denotes the probability density function. We assume that the process has exhibited a change at the time point τ, i.e., the in-control distribution is $f(x;\theta_1)$ and after the change in the parameter, the new distribution is $f(x;\theta_2)$. In this thesis, we focus on identifying the changes in the process mean (μ) and assume that the process variance (σ^2) is unchanged, i.e., we consider the case where $\mu_1 \neq \mu_2$ and $\sigma_1^2 = \sigma_2^2 = \sigma^2$ and the three parameters μ_1, μ_2, and σ^2 are unknown. Moreover, we assume that there is only one change point. In the next section, we will briefly discuss the approaches proposed by Hawkins et al. (2003) and Chen et al. (2006).

2.1 Hawkins Method

Hawkins et al. (2003) developed the change point model for process mean assuming that the observations are normally distributed. Considering each data point as a change point, they perform a series of two-sample t tests for the data sequences before and after the change point (see Sen and Srivastava (1975), Hawkins (1977), and Worsley (1979)). The data point with the most significant test statistic is identified as the candidate change point.

Let $X_1, X_2, \ldots, X_n$ be independently and normally distributed samples of size

n. There are $n-1$ potential change points. For a particular change point j where $1 \leq j \leq n-1$, compute the two-sample t-statistic as

$$T_{jn} = \sqrt{\frac{j(n-j)}{n}} \left[\frac{\bar{X}_{jn} - \bar{X}^*_{jn}}{\hat{\sigma}_{jn}} \right],$$

where $\bar{X}_{jn} = \frac{\sum_{i=1}^{j} X_i}{j}$ is the mean of the first j observations, $\bar{X}^*_{jn} = \frac{\sum_{i=j+1}^{n} X_i}{(n-j)}$ is the mean of the remaining $n-j$ observations, and

$$\hat{\sigma}_{jn} = \sqrt{\left[\sum_{i=1}^{j} (X_i - \bar{X}_{jn})^2 + \sum_{i=j+1}^{n} (X_i - \bar{X}^*_{jn})^2 \right] / (n-2)}$$

is the pooled sample standard deviation. $\bar{X}_{jn}$ and $\bar{X}^*_{jn}$ are the maximum likelihood estimators of μ_1 and μ_2 respectively, under the assumption that there is only one change point j, and $\hat{\sigma}^2_{jn}$ is the estimator for σ^2. If there is no change in the process mean at a chosen time point j, T_{jn} should follow a Student's t-distribution with $n-2$ degrees of freedom.

To identify the most appropriate change point, first we find $T_{\max,n}$, the maximum of $|T_{jn}|$ for all j $(1 \leq j \leq n-1)$. The j which corresponds to $T_{\max,n}$ is the maximum likelihood estimate of the change point τ if $T_{\max,n} > h_n$, where h_n is the critical value. For a given significance level α, we use the Bonferroni inequality to find h_n as

$$\begin{aligned} \Pr[T_{\max,n} > h_n] &\leq \sum_{j=1}^{n-1} \Pr[|T_{jn}| > h_n] \\ &= (n-1)\Pr[|t_{n-2}| > h_n] \end{aligned}$$

where t_{n-2} is the Student's t quantile for $n-2$ degrees of freedom. For large n, this inequality is conservative. For this reason, Hawkins et al. (2003) conducted a large number of simulations to find h_n for different sample sizes and confidence levels $1-\alpha$; these are tabulated in Hawkins et al. (2003). Following this approach, whenever a new observation X_{n+1} is added to the data set, we compute $T_{\max,n+1}$ and compare it with h_{n+1}. If there is a signal, we identify that particular j as the change point. For a particular $n \geq 11$ and hazard rate α, define

$$h_{n,\alpha} \approx h_{10,\alpha}\left(0.677 + 0.019\log(\alpha) + \frac{1-0.115\log(\alpha)}{n-6}\right)$$

To simplify the computational algorithm Hawkins et al. (2003) suggested a recursive procedure. First compute $S_{n-1} = \sum_{i=1}^{n-1} X_i$ and $W_{n-1} = \sum_{i=1}^{n-1}(X_i - \bar{X}_{n-1})^2$, where $\bar{X}_{n-1} = S_{n-1}/n-1$. When a new observation X_n is added to the data, $S_n = S_{n-1} + X_n$ and $W_n = W_{n-1} + ((n-1)X_n - S_{n-1})^2/[n(n-1)]$. Instead of computing T_{jn} for each $1 \leq j \leq n-1$, compute T_{jn}^2 as follows:

$$E_{jn} = (nS_j - jS_n)^2/[nj(n-j)]$$

The analysis of variance identity is given by $V_{jn} = W_n - E_{jn}$, leading to $T_{jn}^2 = (n-2)E_{jn}/(W_n - E_{jn})$. Then compute $T_{\max,n}^2$, the maximum of T_{jn}^2, $j = 1, 2, \ldots, n-1$. If $T_{\max,n}^2 > h_n^2$, the corresponding j is the change point. The same j which maximizes the T_{jn}^2 also maximizes E_{jn}. Therefore, we compute only $E_{\max,n}$, then proceed to

$T^2_{\max,n}$. Hawkins and Zamba (2005) suggested a method for identifying the change point due to a shift in the mean or the variance, which is not considered here since our main focus is on a shift in the process mean.

2.2 Modified Information Criterion

The modified information criterion (MIC) proposed by Chen et al. (2006) is a likelihood-based method for the identification of the change point in regular models. This criterion is a modification of the BIC. They introduced a new penalty function based on the complexity of the regular model. The modeling of the change point problem under MIC is as follows:

Assuming that a change in the process mean occurred at a data point j, we wish to test $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 \neq \mu_2$. For a regular model with log-likelihood function $\ell_n(\mu)$, the Akaike information criterion (AIC) is defined to be

$$AIC = -2\ell_n(\hat{\mu}) + 2\dim(\hat{\mu})$$

and the BIC is

$$BIC = -2\ell_n(\hat{\mu}) + \dim(\hat{\mu})\log(n)$$

where $\hat{\mu}$ is the maximum likelihood estimate of μ and $\dim(X)$ is the dimension of X.

The log-likelihood function for the change point problem is

$$\ell_n(\mu_1, \mu_2, j) = \sum_{i=1}^{j} \log f(x_i; \mu_1) + \sum_{i=j+1}^{n} \log f(x_i; \mu_2)$$

where $1 \leq j \leq n-1$ and $f(.)$ is the probability density function. The BIC for the change point problem becomes

$$BIC(j) = -2\ell_n(\hat{\mu}_{1j}, \hat{\mu}_{2j}, j) + [2\dim(\hat{\mu}_{1j}) + 1]\log(n)$$

where $\hat{\mu}_{1j}$ and $\hat{\mu}_{2j}$ maximize $\ell_n(\mu_1, \mu_2, j)$ for a given data point j. We can rewrite BIC in terms of the complexity(.) as

$$BIC(j) = -2\ell_n(\hat{\mu}_{1j}, \hat{\mu}_{2j}, j) + \text{complexity}(\hat{\mu}_{1j}, \hat{\mu}_{2j}, j)\log(n)$$

In the context of the change point problem, $\text{complexity}(\hat{\mu}_{1j}, \hat{\mu}_{2j}, j) \neq 2\dim(\hat{\mu}_{1j})+1$ because when j is near 1 or n, either μ_1 or μ_2 becomes uninformative. Therefore, Chen et al. (2006) defined a new complexity(.) for change point problems as

$$\text{complexity}(\hat{\mu}_{1j}, \hat{\mu}_{2j}, j) = \left[2\dim(\hat{\mu}_{1j}) + \left(\frac{2j}{n} - 1\right)^2\right] + \text{constant}$$

The modified information criterion for a candidate change point j is defined to be

$$MIC(j) = -2\ell_n(\hat{\mu}_{1j}, \hat{\mu}_{2j}, j) + \left[2\dim(\hat{\mu}_{1j}) + \left(\frac{2j}{n} - 1\right)^2\right]\log(n), \quad j = 1, 2, \ldots, n-1$$

Under the null model (i.e., there is no change point), define

$$MIC(n) = -2\ell_n(\hat{\mu}, \hat{\mu}, n) + \dim(\hat{\mu})\log(n)$$

where $\hat{\mu}$ maximizes $\ell_n(\hat{\mu}, \hat{\mu}, n)$. We can select the model with a change point if $MIC(n) > \min_{1 \le j \le n-1} MIC(j)$ and estimate the change point by $\hat{\tau}$ such that

$$MIC(\hat{\tau}) = \min_{1 \le j \le n-1} MIC(j)$$

Let

$$S_n = MIC(n) - \min_{1 \le j \le n-1} MIC(j) + \dim(\mu) \log(n) \tag{2.1}$$

In this context, $dim(\mu) = dim(\hat{\mu}) = dim(\hat{\mu}_{1j}) = d$. Under H_0, $S_n \sim \chi^2_d$ in distribution as $n \to \infty$. If $S_n > \chi^2_{d,1-\alpha}$, where α is the level of significance, then the corresponding data point is identified as the change point. Chen et al. (2006) showed the following properties for S_n:

1. If there is a change at τ, τ/n has limit in $(0, 1)$ as $n \to \infty$, then $S_n \to \infty$ in probability.

2. The estimator $\hat{\tau}$ has a best convergence rate of $O_p(1)$, and via a random walk it has a limiting distribution too.

The Hawkins method is developed based on the assumption that the quality characteristic is normally distributed, whereas MIC is based on the likelihood ratio test, assuming a parametric distribution of the quality characteristic. Thus, both methods need a well-defined parametric distributional form for the response of interest.

As discussed in the Introduction, in many situations, the parametric distribution of the quality characteristic is not well defined, but we may be able to define a set of estimating equations for the parameters of interest. It is therefore necessary to have a distribution-free method for the situations where the distribution of the observations is not easy to define. We propose an EL-based information criterion (ELIC) to identify the change points when the distribution of the observations is not known.

Chapter 3

Empirical-Likelihood-Based Information Criterion

Before introducing our EL-based information criterion (ELIC) for the change point problem, we briefly explain the basic concepts of empirical likelihood.

3.1 Empirical Likelihood

Likelihood methods for regular parametric models have many good properties. However, there is a significant risk of model misspecification. For example, if the data are collected from a process which follows a Cauchy distribution, but we use a

normal model for the statistical analysis, then our conclusions may not be correct. By the central limit theorem, if the data have a finite variance, then the sample average follows the normal distribution and we can make valid inferences asymptotically. However, efficiency may be compromised. For example, assume that the data follow $N(\mu, \sigma^2)$, and we define the unbiased estimator of σ^2 as $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$, where $\bar{X}$ is the sample mean. When $n \to \infty$, s^2 converges to σ^2. The variance of s^2 is $\sigma^4(\frac{2}{n-1} + \frac{\kappa}{n})$, and by the normality assumption $\mathrm{Var}(s^2) = \frac{2\sigma^4}{n-1}$, since the kurtosis $\kappa = 0$. If κ of the original distribution is non-zero, then the normality assumption will not lead to a good estimate for $n\mathrm{Var}(s^2)$. It will significantly affect the confidence interval of σ^2. Except for the normality assumption, it is not easy to find a distributional assumption that fits the data perfectly and allows both the skewness and kurtosis to vary freely. To avoid this risk of misspecification, nonparametric methods can be used. Empirical likelihood (EL) is a systematic nonparametric approach to statistical analysis, introduced by Owen (1988). EL provides a data-determined shape for confidence regions; it can assimilate known constraints on parameters and adjust for biased sampling schemes. In this thesis, we define EL based on a set of estimating equations for the parameters of interest.

Let $X_1, X_2, \ldots, X_n$ be a set of iid observations having a common distribution function F. We wish to make statistical inference without placing a restrictive assumption

on F. The empirical distribution $F_n(x)$ is a good estimator of F and we can view it as a nonparametric maximum likelihood estimate of F. Let

$$F(x_i) - F(x_i-) = P(X_i = x_i), \ \ i = 1, 2, \ldots, n$$

where $F(x_i-) = P(X_i < x_i)$ and the x_i's are the realization of X_i. If all the x_i's are disjoint, the likelihood function becomes

$$\begin{aligned} L_n(F) &= \prod_{i=1}^{n} \{F(x_i) - F(x_i-)\} \\ &= \prod_{i=1}^{n} P(X_i = x_i) \\ &= \prod_{i=1}^{n} p_i \end{aligned}$$

Clearly $p_i \geq 0, \ i = 1, 2, \ldots, n$, and $\sum_{i=1}^{n} p_i = 1$. The nonparametric empirical log-likelihood function of F is given by

$$l_n(F) = \sum_{i=1}^{n} \log p_i$$

$L_n(F)$ cannot be a likelihood function in terms of p_i if there are ties in the data. To make $L_n(F)$ a likelihood function, we can add a set of small and independent continuous noises to the x_i's. This likelihood function is maximized when F is equal to the empirical distribution function, F_n. Since $L_n(F_n) > L_n(F)$ for any $F \neq F_n$, it is convenient to standardize $L_n(F)$ by division by $L_n(F_n)$. This convention leads to

the EL ratio

$$\begin{aligned} R_n(F) &= \frac{L_n(F)}{L_n(F_n)} \\ &= \frac{\prod_{i=1}^{n} p_i}{(1/n)^n} \\ &= \prod_{i=1}^{n} (np_i) \end{aligned}$$

and the empirical log-likelihood ratio is

$$r_n(F) = \sum_{i=1}^{n} \log(np_i)$$

3.1.1 Profile Empirical Likelihood

The targeted application of EL is the inference of parameters in the form of some functional of the population distribution, say $\phi = T(F)$. Inferences will be made under the assumption that $F \in \mathcal{F}$, the nonparametric distribution family.

If we have a likelihood value at ϕ, we can make inferences about ϕ by the likelihood approach. For each given value of ϕ, there are many members of F in $\mathcal{F}$ such that $T(F) = \phi$. We must decide which F best represents ϕ. The concept of profile empirical likelihood is to find the F for which EL attains a maximum among the set satisfying $T(F) = \phi$. The profile empirical likelihood is defined to be

$$L_n(\phi) = \sup\{L_n(F) \mid T(F) = \phi;\ \ F \in \mathcal{F}\}$$

This likelihood has similar properties to its parametric counterpart. The inference on ϕ can therefore be based on $L_n(\phi)$.

Since $L_n(\phi) \leq n^{-n} = L_n(F_n)$, it is convenient to standardize $L_n(\phi)$ by defining the ratio function

$$R_n(\phi) = \frac{L_n(\phi)}{L_n(F_n)}$$
$$= n^n L_n(\phi)$$

The empirical log-likelihood ratio function is

$$r_n(\phi) = n \log n + \log L_n(\phi) = n \log n + l_n(\phi)$$

Let $\phi_0 = E(X_1)$ and $Var(X_1) < \infty$. Then

$$-2 \log[R_n(\phi_1)] \xrightarrow{L} \chi^2_d \text{ as } n \to \infty$$

where $d = \dim(X) = \dim(\phi)$ (Owen (1988, 2001)). Using the above result, we obtain the $100(1-\alpha)\%$ confidence interval for ϕ:

$$\{\phi : -2 \log[R_n(\phi_0)] \leq \chi^2_{d,1-\alpha}\}$$

3.1.2 Optimization

Suppose the parameter of interest $\phi = (\phi_1, \phi_2, \ldots, \phi_p)$ is specified by an estimating equation $Eg(x, \phi) = 0$, where $g(.)$ is a real valued function. The profile EL for ϕ is

obtained by maximizing

$$l_n(F) = \sum_{i=1}^{n} \log p_i$$

under the constraints $p_i \geq 0,\ i = 1, 2, \ldots, n,\ \sum_{i=1}^{n} p_i = 1$, and $\sum_{i=1}^{n} p_i g(x_i, \phi) = 0$. To solve this maximization problem, Lagrange multiplier method is very much effective.

Define

$$G(s, \lambda) = \sum_{i=1}^{n} \log p_i + s(\sum_{i=1}^{n} p_i - 1) - n\lambda(\sum_{i=1}^{n} p_i g(x_i, \phi))$$

where s and λ are Lagrange multipliers. By Lagrange Lagrange multiplier method, we need to find the stationary points of $G(s, \lambda)$ with respect to s and λ. We obtain $s = n$ and λ can be obtained by solving $\sum_{i=1}^{n} p_i g(x_i, \phi) = 0$. Using this Lagrange multiplier method, EL is maximized when

$$\hat{p}_i = \frac{1}{n\{1 + \hat{\lambda}^t g(x_i, \phi)\}},\quad i = 1, 2, \ldots, n.$$

Then, the profile EL for ϕ is defined to be

$$l_n(\phi) = -n\log(n) - \sum_{i=1}^{n} \log\{1 + \hat{\lambda}^t g(x_i, \phi)\}$$

and the EL ratio function is

$$r_n(\phi) = -\sum_{i=1}^{n} \log\{1 + \hat{\lambda}^t g(x_i, \phi)\}$$

For a given ϕ, the EL ratio statistic is defined to be

$$W(\phi) = -2r_n(\phi) = 2\sum_{i=1}^{n} \log\{1 + \hat{\lambda}^t g(x_i, \phi)\}$$

This converges to χ^2_d when $n \to \infty$, where d is the dimension of ϕ.

3.2 Two-Sample Problem Using Empirical Likelihood

In the context of the change point problem, we first introduce the two-sample problem using empirical likelihood (see Jing (1995) and Liu, Zou and Zhang (2008)).

Let $X_1, X_2, \ldots, X_{n_1}$ and $Y_1, Y_2, \ldots, Y_{n_2}$ be iid samples from two probability distributions $f(x;\mu)$ and $f(y;\mu-\delta)$. We wish to test the hypothesis

$$H_0 : \delta = \delta_0 = 0 \text{ against } H_1 : \delta \neq \delta_0$$

Define $\theta = n_1/n$, where $n = n_1 + n_2$, and assume that $\theta = n_1/n \to \theta_0 \in (0,1)$ as $n \to \infty$. Let $(p_1, p_2, \ldots, p_{n\theta})$ and $(q_1, q_2, \ldots, q_{n(1-\theta)})$ be probability vectors corresponding to $X_1, X_2, \ldots, X_{n_1}$ and $Y_1, Y_2, \ldots, Y_{n_2}$ respectively such that $\sum_{i=1}^{n_1} p_i = 1$, $\sum_{l=1}^{n_2} q_l = 1$, $p_i \geq 0$; $i = 1, 2, \ldots, n_1$, and $q_l \geq 0$; $l = 1, 2, \ldots, n_2$.

Suppose the mean parameter μ is specified by an estimating equation $Eg(x,\mu) = E(x-\mu) = 0$ corresponding to the first group of observations and $Eh(y,\mu) = E(y-\mu) = 0$ corresponding to the second group under null hypothesis. The profile EL for

μ is obtained by maximizing

$$L_n(F) = \prod_{i=1}^{n_1} p_i \prod_{l=1}^{n_2} q_l$$

and the corresponding empirical log-likelihood is given by

$$l_n(F) = \sum_{i=1}^{n_1} \log p_i + \sum_{l=1}^{n_2} \log q_l$$

under the constraints

$$p_i \geq 0,\ i = 1, 2, \ldots, n_1, \quad q_l \geq 0,\ l = 1, 2, \ldots, n_2$$

$$\sum_{i=1}^{n_1} p_i = 1, \quad \sum_{l=1}^{n_2} q_l = 1$$

$$\sum_{i=1}^{n_1} p_i g(x_i, \mu) = 0, \quad \sum_{l=1}^{n_2} q_l h(y_l, \mu) = 0$$

Using the Lagrange multiplier method (as discussed in Section 3.1.2), EL is maximized when

$$\hat{p}_i = \frac{1}{n\theta\{1 + \theta^{-1}\hat{\lambda}_1^t g(x_i, \mu)\}}, \quad i = 1, 2, \ldots, n_1$$

$$\hat{q}_l = \frac{1}{n(1-\theta)\{1 + (1-\theta)^{-1}\hat{\lambda}_2^t h(y_l, \mu)\}}, \quad l = 1, 2, \ldots, n_2.$$

The Lagrange multipliers λ_1 and λ_2 and the mean μ are the solution of the following equations:

$$\left.\begin{aligned} &\sum_{i=1}^{n_1} p_i g(x_i, \mu) = 0 \\ &\sum_{l=1}^{n_2} q_l h(y_l, \mu) = 0 \\ &\lambda_1 + \lambda_2 = 0 \end{aligned}\right\} \tag{3.1}$$

Then the profile EL for μ is given by

$$\begin{aligned} l_{EL}(\mu) = &- n\theta \log(n\theta) - \sum_{i=1}^{n_1} \log\{1 + \theta^{-1}\hat{\lambda}_1^t g(x_i, \mu)\} \\ &- n(1-\theta)\log\{n(1-\theta)\} - \sum_{i=1}^{n_2} \log\{1 + (1-\theta)^{-1}\hat{\lambda}_2^t h(y_i, \mu)\} \end{aligned} \tag{3.2}$$

For a given μ, the EL ratio statistic is defined to be

$$W(\mu) = 2[\sum_{i=1}^{n_1} \log\{1 + \theta^{-1}\hat{\lambda}_1^t g(x_i, \mu)\} + \sum_{i=1}^{n_2} \log\{1 + (1-\theta)^{-1}\hat{\lambda}_2^t h(y_i, \mu)\}] \tag{3.3}$$

This converges to χ_d^2 when $n \to \infty$, where d is the dimension of μ.

Based on this two-sample problem using EL, we propose an EL-based information criterion (ELIC) for the change point problem.

3.3 EL-Based Information Criterion

Our method is a modification of MIC, where we replace the parametric likelihood by the empirical likelihood in (2.1) based on a set of estimating equations. We are interested in the change in the mean μ. The sample $X_1, X_2, \ldots, X_n$ can be considered to be two groups divided by the data point at j. The first group contains n_1 observations $(1, 2, \ldots, j)$ and the second group contains n_2 observations $(j+1, j+2, \ldots, n)$. To derive the EL ratio statistic for ELIC, we adopt the procedure discussed in the previous section with $\theta = j/n$. Let $(p_1, p_2, \ldots, p_j)$ and $(q_{j+1}, q_{j+2}, \ldots, q_n)$ be probability

vectors corresponding to $X_1, X_2, \ldots, X_{n_1}$ and $X_1, X_2, \ldots, X_{n_2}$ respectively. The main difference is in the estimating equations because the mean μ is specified by an estimating equation $Eg(x_i, \mu) = E(x_i - \mu) = 0,\ i = 1, 2, \ldots, j$, corresponding to the first group of n_1 observations and $Eh(x_l, \mu) = E(x_l - \mu) = 0,\ l = j+1, j+2, \ldots, n$, corresponding to the second group of n_2 observations under null hypothesis, $H_0 : \delta = 0$. The profile EL for μ is found by maximizing

$$L_n(F) = \prod_{i=1}^{n_1} p_i \prod_{l=1}^{n_2} q_l$$

under the same constraints with the necessary changes in the estimating equations. Using the Lagrange multiplier method, EL is maximized for

$$\hat{p}_i = \frac{1}{n\theta\{1 + \theta^{-1}\hat{\lambda}_1^{\,t} g(x_i, \mu)\}},\quad i = 1, 2, \ldots, n_1$$

$$\hat{q}_l = \frac{1}{n(1-\theta)\{1 + (1-\theta)^{-1}\hat{\lambda}_2^{\,t} h(x_l, \mu)\}},\quad l = 1, 2, \ldots, n_2.$$

The Lagrange multipliers λ_1 and λ_2 and the mean μ are the solution of (3.1) with the new estimating equations. Based on this, the new EL ratio function for the change point at j, which is a function of j, is defined to be

$$W_j(\mu) = 2[\sum_{i=1}^{j} \log\{1 + \theta^{-1}\hat{\lambda}_1^{\,t} g(x_i, \mu)\} + \sum_{l=j+1}^{n} \log\{1 + (1-\theta)^{-1}\hat{\lambda}_2^{\,t} h(x_l, \mu)\}] \quad (3.4)$$

By considering the no-change-point case as well as a change point at j, we define the ELIC for each data point j using the EL ratio statistic as:

$$ELIC(j) = W_j(\mu) - (\frac{2j}{n} - 1)^2 \log(n) \quad j = 2, \ldots, n-2 \tag{3.5}$$

where $W_j(\mu)$ is defined in (3.4).

Similarly to the MIC method, it can be proved that when there is no change point,

$$ELIC(\tau) = \max_{2 \le j \le n-2} ELIC(j) \sim \chi^2_d, \quad as \quad n \to \infty.$$

If $ELIC(\tau) > \chi^2_{d,1-\alpha}$, then the corresponding j is the change point.

We know that the parametric likelihood ratio statistic has a χ^2 limiting distribution; this is one of the major properties used to show that S_n (2.1) for MIC follows a χ^2_d distribution. Similarly, Owen (1988, 2001) proved that the EL ratio statistic also has a χ^2 limiting distribution. Variyath (2006) and Variyath et al. (2010) proved that the EL-based information criterion for variable selection also has a χ^2 approximation. We have modified MIC by replacing the parametric likelihood with the EL based on a set of estimating equations, and it can easily be shown that ELIC follows the χ^2 with some modifications in the MIC proof. However, our simulation studies in Section 3.6 for univariate data show that the distributions of MIC and ELIC cannot be approximated by a χ^2 distribution when we consider a sample size ($n = 200$) feasible for industrial situations. For $n = 200$, we derived an empirical distribution for ELIC and

MIC based on 10000 simulations. If the false-alarm rate is 0.05, the 95^{th} percentile should be close to $\chi^2_{1,0.95} = 3.841459$. However, for $n = 200$ our quantiles for MIC and ELIC are 7.821 and 8.138 respectively. Note that for larger sample sizes (say n=1000), the 95^{th} quantile of ELIC is 3.9157, which is close to $\chi^2_{1,0.95}$. However, in industry, a sample size of $n = 1000$ is large and therefore expensive. Therefore, the properties and distribution of the statistics MIC and ELIC under the null hypothesis that there is no change point are investigated using its empirical distribution based on large number of simulations (details are given in Section 3.6.1). Since we did not use the χ^2 approximation, the theoretical proof is omitted here.

3.4 Technical Problem and Adjusted Empirical Likelihood

The computation of ELIC may face some technical issues in situations where n_1 or n_2 is small. This is mainly because of the nonexistence of solutions to (3.1) when 0 is not an inner point of the convex hull of the corresponding estimating function values. This problem occurs when we compute ELIC for data points close to the ends of the data set, because one sample has much fewer data points than the other. When the number of data points in one sample is small, the convex hull of these points

may not contain the current estimate of the mean. This leads to the nonexistence of solutions to EL, and the algorithm for ELIC will break down. To overcome this problem, we propose using adjusted EL (AEL) (Chen et al. (2008) and Variyath et al. (2010)). Let $g_i = g(x_i, \mu)$, $i = 1, 2, \ldots, n_1$, and $h_l = h(x_l, \mu)$, $l = 1, 2, \ldots, n_2$, be the estimating function values. In AEL, we add additional estimating function points $g_{n_1+1} = -\bar{g}_{n_1} * a_{n_1}$ and $h_{n_2+1} = -\bar{h}_{n_2} * a_{n_2}$ for some positive constants a_{n_1} and a_{n_2}, where $\bar{g}_{n_1}$ and $\bar{h}_{n_2}$ are the means of the g_i's and h_i's respectively. This ensures that solutions to (3.1) always exist. Then the ELIC formulation will be modified as follows. The profile EL for μ is obtained by maximizing

$$L_n^*(F) = \prod_{i=1}^{n_1+1} p_i \prod_{l=1}^{n_2+1} q_l$$

and the corresponding empirical log-likelihood is given by

$$l_n^*(F) = \sum_{i=1}^{n_1+1} \log p_i + \sum_{l=1}^{n_2+1} \log q_l$$

under the constraints

$$p_i \geq 0,\ i = 1, 2, \ldots, n_1 + 1, \quad q_l \geq 0,\ l = 1, 2, \ldots, n_2 + 1$$

$$\sum_{i=1}^{n_1+1} p_i = 1, \quad \sum_{l=1}^{n_2+1} q_l = 1$$

$$\sum_{i=1}^{n_1+1} p_i g_i = 0, \quad \sum_{l=1}^{n_2+1} q_l h_l = 0$$

Using the Lagrange multiplier method, EL is maximized at

$$\hat{p}_i = \frac{1}{n\theta^*\{1 + \theta^{*-1}\hat{\lambda}_1^t g_i\}}, \quad i = 1, 2, \ldots, n_1 + 1$$

$$\hat{q}_l = \frac{1}{n(1-\theta^*)\{1 + (1-\theta^*)^{-1}\hat{\lambda}_2^t h_l\}}, \quad l = 1, 2, \ldots, n_2 + 1$$

where $\theta^* = (n_1+1)/(n+2)$. The Lagrange multipliers λ_1 and λ_2 and the mean μ are the solution of the following equations:

$$\left.\begin{aligned} \sum_{i=1}^{n_1+1} p_i g_i &= 0 \\ \sum_{l=1}^{n_2+1} q_l h_l &= 0 \\ \lambda_1 + \lambda_2 &= 0 \end{aligned}\right\} \tag{3.6}$$

Then the profile EL for μ is given by

$$\begin{aligned} l^*_{EL}(\mu) = &- n\theta^* \log(n\theta^*) - \sum_{i=1}^{n_1+1} \log\{1 + \theta^{*-1}\hat{\lambda}_1^t g_i\} \\ &- n(1-\theta^*)\log(n(1-\theta^*)) - \sum_{l=1}^{n_2+1} \log\{1 + (1-\theta^*)^{-1}\hat{\lambda}_2^t h_l\} \end{aligned} \tag{3.7}$$

The adjusted EL ratio function is

$$W_j^*(\mu) = 2[\sum_{i=1}^{n_1+1} \log\{1 + \theta^{*-1}\hat{\lambda}_1^t g_i\} + \sum_{l=1}^{n_2+1} \log\{1 + (1-\theta^*)^{-1}\hat{\lambda}_2^t h_l\}] \tag{3.8}$$

This converges to χ^2_d when $n \to \infty$ (Chen et al. (2008)). We use $W_j^*(\mu)$ instead of $W_j(\mu)$ in (3.5) for ELIC for change point problems as

$$ELIC(j) = W_j^*(\mu) - (\frac{2j}{n} - 1)^2 \log(n) \quad j = 2, \ldots, n-2 \tag{3.9}$$

For simplicity, hereafter ELIC refers to the adjusted ELIC. It can be proved that

$$ELIC(\tau) = \max_{2 \le j \le n-2} ELIC(j) \sim \chi^2_d, \quad as \quad n \to \infty. \tag{3.10}$$

We use $a_{n_1} = -\log(n_1)/2$ and $a_{n_2} = -\log(n_2)/2$ in our simulation studies. For more details on AEL, refer to Chen et al. (2008) and Variyath et al. (2010).

3.5 Computational Algorithm

We used the modified Newton-Raphson algorithm of Chen, Sitter and Wu (2002) to compute the EL ratio statistic given in (3.8). A candidate change-point j divides the sample data into two groups. The first j observations form the first group (total of n_1 observations) and the remaining $n-j$ observations form the second group (total of n_2 observations). For each data point j, we compute

$$W_j^*(\mu) = 2[\sum_{i=1}^{n_1+1} \log\{1 + \theta^{*-1}\hat{\lambda}_1^t g_i\} + \sum_{l=1}^{n_2+1} \log\{1 + (1-\theta^*)^{-1}\hat{\lambda}_2^t h_l\}]$$

as follows:

1. Initialize: $k = 1$, $\lambda_1 = 0$, $\lambda_2 = 0$, $\mu = \bar{x}$, and the step-down parameter $\gamma = 1$.

2. In the k^{th} iteration, compute the $g_i's$, $i = 1, 2, \ldots, n_1$ and $h_l's$, $l = 1, 2, \ldots, n_2$. Here $g_i = g(x_i, \mu) = (x_i - \mu)$ and $h_l = h(x_l, \mu) = (x_l - \mu)$. Then add the

additional points $g_{n_1+1} = -\bar{g}_{n_1} * a_{n_1}$ and $h_{n_2+1} = -\bar{h}_{n_2} * a_{n_2}$ to the first group and second group respectively. Represent (3.6) as

$$A = \begin{bmatrix} A_1 = \sum_{i=1}^{n_1} p_i g_i + p_{n_1+1} g_{n_1+1} = 0 \\ A_2 = \sum_{l=1}^{n_2} q_l h_l + q_{n_2+1} h_{n_2+1} = 0 \\ A_3 = \lambda_1 + \lambda_2 = 0 \end{bmatrix}$$

Compute the derivatives of A with respect to $\beta = (\lambda_1, \lambda_2, \mu)$ as

$$D = \begin{bmatrix} \frac{\partial A_1}{\partial \lambda_1} & \frac{\partial A_1}{\partial \lambda_2} & \frac{\partial A_1}{\partial \mu} \\ \frac{\partial A_2}{\partial \lambda_1} & \frac{\partial A_2}{\partial \lambda_2} & \frac{\partial A_2}{\partial \mu} \\ \frac{\partial A_3}{\partial \lambda_1} & \frac{\partial A_3}{\partial \lambda_2} & \frac{\partial A_3}{\partial \mu} \end{bmatrix}$$

3. Using the Newton-Raphson procedure, compute the iterative estimates of $\beta = (\lambda_1, \lambda_2, \mu)$ as

$$\beta^{k+1} = \beta^k - \gamma D^{-1} A$$

4. Check the convergence: if $|\beta^k - \beta^{k+1}| <= 10^{-5}$, then go to step 6; otherwise continue.

5. Feasibility checking: if either of the following occurred, set $\gamma = \gamma/2$, then return to step 3.

   - $\min(p_i) < 0, \quad i = 1, 2, \ldots, n_1 + 1$

- $\min(q_l) < 0, \quad l = 1, 2, \ldots, n_2 + 1$

Otherwise, set $k = k + 1$ and go to step 2.

6. Compute $W_j^*(\mu)$ from (3.8) and find $ELIC(j)$ from (3.9).

Repeat steps 1 to 6 to compute $ELIC(j)$, $j = 2, 3, \ldots, n - 2$, and find the maximum, $ELIC(\tau)$.

3.6 Performance Analysis

In the performance analysis, we compare ELIC only with MIC. We do not consider the Hawkins method for two reasons. First, the Hawkins method is based on the assumption that the quality characteristic is normally distributed, which limits its application. Second, our estimated signal probability based on the Hawkins method for the change in the process mean is high compared to the significance level, α, when there is no change point. This also results in a high level of signal probability when the process mean shifts. Therefore, we omitted the Hawkins method from our comparison.

3.6.1 Empirical Distribution of MIC and ELIC

To assess the performance of our method, we conducted a series of simulation studies in univariate data. The identification of the change point in mean of a given data set is the major focus of the simulations. Our initial simulations indicated that when there is no change point, the χ^2 approximation for ELIC and MIC does not fit well even for a modest sample size of $n = 200$, which is a reasonable sample size for change point problems (see Chen et al. (2006)). We therefore decided to construct empirical distributions for ELIC and MIC to find the critical values for $n = 200$, when there is no change point. We consider data from the standard normal distribution and exponential distribution with mean $\mu = 1$ when constructing the empirical distribution for ELIC and MIC, which is inverted to give the critical values. The critical values for ELIC and MIC based on 100000 simulations for normal and exponential data are similar, so we pooled them. The pooled critical values found for $n = 200$ and for different false-alarm rates, $\alpha = 0.1$, 0.05, 0.01, and 0.0027, are given in Table 3.1.

Method	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.0027$
MIC	6.253726	7.821014	11.383118	14.170007
ELIC	6.444047	8.137872	12.423956	16.647301

Table 3.1: Critical Values for MIC and ELIC

3.6.2 Univariate Normal Data

To assess the ability of each method to identify the change point, we generated data sets with $n = 200$ and a change in the mean from the 101^{st} ($\tau = 100$) data point onwards. The changes in the mean considered for the standard normal distribution are 0, 0.2, 0.4, 0.6, 0.8, and 1.0. For example, if $\tau = 100$ and the shift is 0.2, we generated the first 100 observations from $N(0, 1)$ and the remaining 100 observations from $N(0.2, 1)$.

For each data set generated, we computed $ELIC(\tau)$ using (3.10) and S_n using (2.1). A change in the mean is detected if these values are greater than the corresponding critical values in Table 3.1 with the appropriate false-alarm rates (α). The corresponding data point is the change point for both ELIC and MIC. We repeated the simulations 5000 times and estimated the signal probability as the proportion of simulations where we identified a change point for each shift. When there is no change in the mean, we expect that the signal probability should be equal to the de-

sired significance level, and when there is a change in the mean, the signal probability should be high. A high signal probability for a method reflects its ability to detect the change in the mean. We considered the significance levels α =0.05 and 0.01. The estimated signal probabilities are given in Figs. 3.1 and 3.2 for different change-point scenarios with normally distributed simulated data.

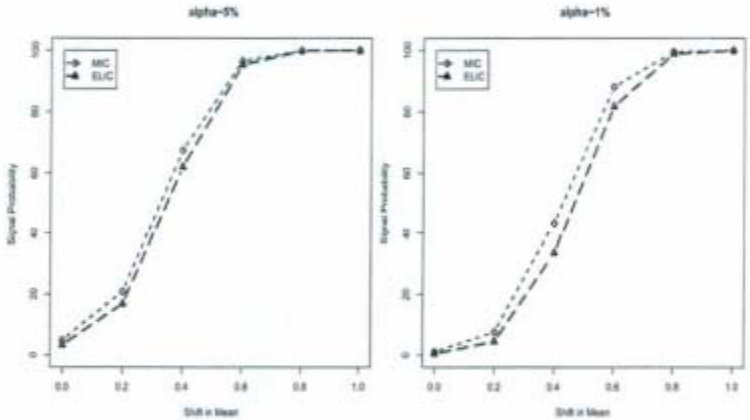

Figure 3.1: Signal probability for N(0,1) data with change point at 100 ($\tau = 100$)

Figure 3.1 depicts the signal probability for the identification of the change point for the normally distributed data, where the mean shifted from the 101st data point onwards (i.e., the change point is at 100). From Fig. 3.1, we see that when the process has an upward shift in the mean, the signal probability increases similarly for both ELIC and MIC. Therefore, ELIC is as good as MIC. Note that in ELIC, we did

not assume any distribution for the data. i.e., the nonparametric method is working as well as the parametric method.

We also considered a change in the mean from the 151st ($\tau = 150$) data point onwards. For example, if $\tau = 150$ and the shift is 0.8, we generated the first 150 observations from $N(0, 1)$ and the remaining 50 observations from $N(0.8, 1)$.

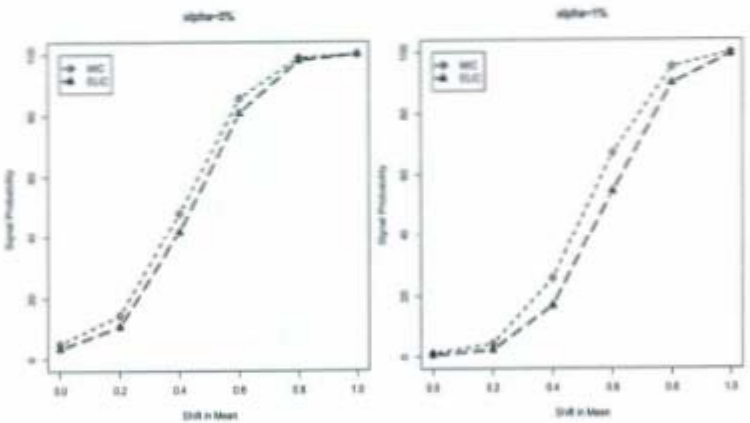

Figure 3.2: Signal probability for N(0,1) data with change point at 150 ($\tau = 150$)

Figure 3.2 shows the signal probability for the identification of the change point for the normally distributed data, where the change point is at 150. We see a similar trend to that in Fig. 3.1, and this supports our earlier claim about the performance of ELIC.

3.6.3 Univariate Exponential Data

To assess the ability of each method to identify the change point in non-normal data, we consider exponentially distributed data sets with $n = 200$ and a change in the mean from the 101^{st} ($\tau = 100$) data point onwards. We considered a downwards change in the mean for the exponential distribution with $\mu = 1$, and the mean shifts are $0, -0.1, -0.2, -0.3, -0.4$, and -0.5 (i.e., the mean changes from 1 to 0.5). For example, if $\tau = 100$ and the shift is -0.2, we generated the first 100 observations from exp(1) and the remaining 100 observations from exp(0.8). Here, exp(μ) means the exponential distribution with mean μ. We followed the procedure discussed previously to identify the change point for both MIC and ELIC. We also consider a change in the mean at the 151^{st} ($\tau = 150$) data point.

Figures 3.3 and 3.4 show the signal probability for the identification of the change point for the exponential data, where the mean shifted from the 101^{st} and 151^{st} data points onwards. We also consider a special case for MIC by assuming a likelihood with a normal distribution when the data are from the exponential distribution. This type of misspecification is common since the distribution of the quality characteristic is unknown and a practitioner will generally assume a normal distribution. We see from Fig. 3.3 that the signal probabilities for ELIC and MIC are similar, indicating that ELIC performs as well as the parametric method, when the distribution is known. It

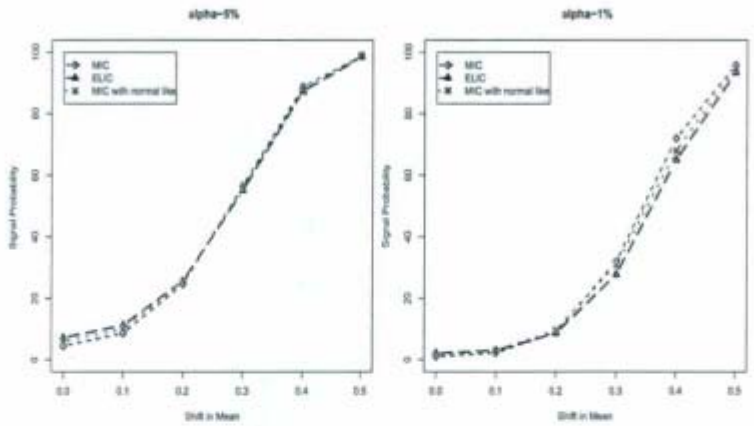

Figure 3.3: Signal probability for exp(1) data with change point at 100 ($\tau = 100$)

is interesting to note that the signal probability for MIC is almost the same whether the likelihood is based on normal or exponential distribution. It can easily be shown that when $n_1 = n_2$ and the true distribution is exponential, the likelihood ratio statistics based on the normal and exponential distributions are the same under the null hypothesis. When there is a shift in mean, the likelihood statistics will differ only by $\log\{1 + \frac{\delta^2}{2\bar{x}(\bar{x}+\delta)}\}$, where $\bar{x}$ is the process mean before shift and δ is the quantum of shift. If $\delta < 1$, then $|\frac{\delta^2}{2\bar{x}(\bar{x}+\delta)}| \leq 1$. Using the Taylor series expansion, $\log\{1 + \frac{\delta^2}{2\bar{x}(\bar{x}+\delta)}\}$ can be approximated as $\frac{\delta^2}{2\bar{x}(\bar{x}+\delta)}$, which is negligibly small. This is why MIC has similar performance with normal and exponential likelihoods. For the exponential data, when $n_1 > n_2$, the likelihood ratios are different when we use the normal

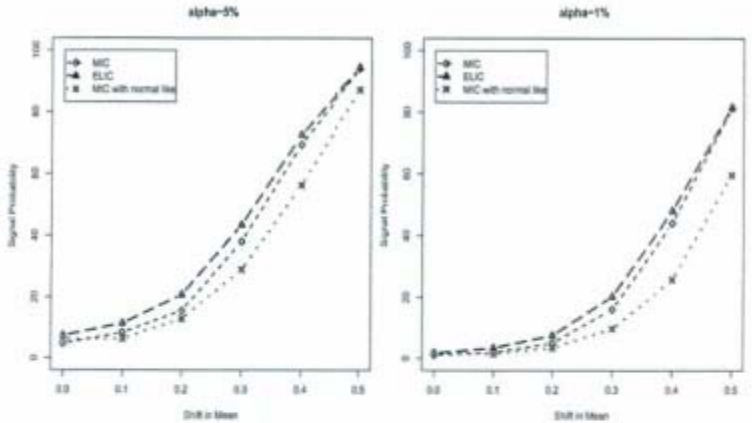

Figure 3.4: Signal probability for exp(1) data with change point at 150 ($\tau = 150$)

distribution (approximation) and the exponential distribution. In Fig. 3.4 (where in reality $n_1 >> n_2$), MIC with the normal distribution (approximation) fails to identify the change point compared to MIC with the true distribution, and ELIC performs as well as MIC with the true distribution. Thus, when we misspecify the distribution of the data, MIC does not perform better than ELIC. Note that in practical situations, we expect the change point to be close to the end of the data set rather than in the middle, so it is always preferable to use ELIC.

The simulation studies lead to two important conclusions. First, when the distribution of the response of interest is known, ELIC and MIC are equally good at identifying the change point. When the distribution of the response is not known,

which may lead to the use of an approximate distribution in MIC, ELIC performs better than MIC. This gives ELIC a clear advantage, since we do not need to specify the parametric distribution.

Chapter 4

EM Test

4.1 Bayesian Approach for Change Point Problem

Bansal et al. (2008) suggested a Bayesian approach for the identification of the change point when the data follow an exponential-family distribution. We briefly review this paper using their notations. Let $X = X_1, X_2, \ldots, X_n$ be a sequence of independent random variables with probability densities

$$f(x_i; \theta_1, \rho) \quad \text{for} \quad i = 1, 2, 3, \ldots, \tau$$

$$f(x_i; \theta_2, \rho) \quad \text{for} \quad i = \tau + 1, \ldots, n,$$

where θ_1 and θ_2 are the parameters before and after the unknown change point τ and ρ is a common nuisance parameter. In this semi-Bayesian approach, they considered the

change point to be random with prior probability distribution $\{\omega(\tau), \tau = 1, 2, \ldots, n\}$. In particular, $\omega(n)$ is the probability that no change occurred in the process mean. The rationale for considering τ to be random is that a change in the mean can occur at random as a natural realization of the process. Then the likelihood function can be written as

$$L(\mu_1, \mu_2, \rho|x) = \sum_{\tau=1}^{n} \omega(\tau) L_\tau(\mu_1, \mu_2, \rho|x)$$

where

$$L_\tau(\mu_1, \mu_2, \rho|x) = \begin{cases} \prod_{i=1}^{\tau} f(x_i; \mu_1, \rho) \prod_{j=\tau+1}^{n} f(x_j; \mu_2, \rho), & \text{for } \tau = 1, 2, \ldots, n-1 \\ \prod_{i=1}^{n} f(x_i; \mu_1, \rho), & \text{for } \tau = n \end{cases}$$

If τ is random, the joint density is given by

$$f(x; \mu_1, \mu_2, \rho, \omega) = \sum_{\tau=1}^{n} \omega(\tau) f(x|\tau; \mu_1, \mu_2, \rho)$$

where

$$f(x|\tau; \mu_1, \mu_2, \rho) = \prod_{i=1}^{\tau} f(x_i; \mu_1, \rho) \prod_{j=\tau+1}^{n} f(x_j; \mu_2, \rho), \quad \tau = 1, 2, \ldots, n-1$$

First we consider the nonparametric case of $\omega(\tau)$. Assume that $\omega(\tau) \geq 0$ and $\sum_{\tau=1}^{n} \omega(\tau) = 1$. Let $\hat{\tau}, \hat{\mu}_{1\tau}, \hat{\mu}_{2\tau}$, and $\hat{\rho}_\tau$ be the estimates that maximize $f(x|\tau; \mu_1, \mu_2, \rho)$. Then the maximum likelihood estimator of $\omega(\tau)$ is given by

$$\hat{\omega}(\tau) = \begin{cases} 1 & \text{if } \tau = \hat{\tau} \\ 0 & \text{if } \tau \neq \hat{\tau} \end{cases}$$

Here $\hat{\tau} \neq n$, since $\tau = n$ implies $\mu_1 = \mu_2$ (no change in the mean). Hence, $1 \leq \tau \leq n-1$.

Let p be a preassigned probability that there is no change point. Then the likelihood function is

$$L(\mu_1, \mu_2, \rho, q : x) = (1-p)\sum_{\tau=1}^{n-1} q(\tau) f(x|\tau; \mu_1, \mu_2, \rho) + pf(x|n; \mu_1, \rho) \tag{4.1}$$

Now the estimation of the change point becomes the estimation of the probability distribution of $q(\tau)$. The maximum likelihood estimator of $q(.)$ is

$$q(\tau) = \begin{cases} 1 & if \sup_{\mu_1,\mu_2,\rho} L_\tau^p(\mu_1, \mu_2, \rho|x) = \max_{1\leq s\leq n-1} \sup_{\mu_1,\mu_2,\rho} L_s^p(\mu_1, \mu_2, \rho|x) \\ 0, & \text{otherwise} \end{cases}$$

where, for $1 \leq s \leq n-1$,

$$L_s^p(\mu_1, \mu_2, \rho|x) = (1-p) f(x|s; \mu_1, \mu_2, \rho) + pf(x|n; \mu_1, \rho) \tag{4.2}$$

Bansal et al. (2008) used the expectation-maximization (EM) algorithm (Dempster, Laird and Rubin (1977)) to estimate the change point by maximizing $L(\mu_1, \mu_2, \rho, q : x)$. The EM algorithm is an iterative method which alternates between performing an expectation (E) step and a maximization (M) step. An E-step computes the expectation of the log-likelihood evaluated using the current estimate for the latent variables, and a M-step computes the parameters that maximize the

expected log-likelihood found in the E-step. Considering τ $(= 1, 2, \ldots, n)$ as a missing value, define the complete data likelihood as follows:

$$L_c(\mu_1, \mu_2, \rho, q; x, \tau) = \begin{cases} pf(x|n; \mu_1, \rho) & \text{when } \tau = n \\ (1-p)q(\tau)f(x|\tau; \mu_1, \mu_2, \rho) & \text{when } \tau = 1, 2, \ldots, n-1 \end{cases}$$

To avoid complex notation, we denote $f(x|t; \hat{\mu}_1^{(j)} \hat{\mu}_2^{(j)}, \hat{\rho}^{(j)})$ by $\hat{f}^{(j)}(x|t)$ when $t = 1, 2, \ldots, n-1$ and $f(x|n; \hat{\mu}_1^{(j)}, \hat{\rho}^{(j)})$ by $\hat{f}^{(j)}(x|n)$, where $\hat{\mu}_1^{(j)}, \hat{\mu}_2^{(j)}, \hat{\rho}^{(j)}$, and $\hat{q}^{(j)}$ are the estimates from the j^{th} iteration of the M-step. Then the conditional distribution of τ given $x; \hat{\mu}_1^{(j)}, \hat{\mu}_2^{(j)}, \hat{\rho}^{(j)}, \hat{q}^{(j)}$ is given by

$$P(\tau = t|x; \hat{\mu}_1^{(j)}, \hat{\mu}_2^{(j)}, \hat{\rho}^{(j)}, \hat{q}^{(j)}) = \begin{cases} \left[\hat{c}^{(j)}(x)\right]^{-1} p\hat{f}^{(j)}(x|n), & t = n \\ \left[\hat{c}^{(j)}(x)\right]^{-1} (1-p)\hat{q}^{(j)}(t)\hat{f}^{(j)}(x|t), & t = 1, 2, \ldots, n-1 \end{cases}$$

$$= \hat{\omega}^{(j)}(t) \text{ (say)} \tag{4.3}$$

where

$$\hat{c}^{(j)}(x) = (1-p)\sum_{s=1}^{n-1} \hat{q}^{(j)}(s)\hat{f}^{(j)}(x|s) + p\hat{f}^{(j)}(x|n).$$

Remark: We can interpret $\hat{\omega}^{(j)}(n)$ as the empirical posterior probability of no change occurring. The prior probability p of the observed data is updated by $\hat{\omega}^{(j)}(n)$. If $\hat{\omega}^{(j)}(n) > d$, the appropriate constant, then we can not reject the hypothesis of no change point. This is equivalent to $\sum_{s=1}^{n-1} \hat{q}(s)\hat{f}(x|s)/\hat{f}(x|n) < \frac{p}{1-p}(d^{-1} - 1)$.

For a single-change-point model and for given $\hat{\mu}_1^{(j)}, \hat{\mu}_2^{(j)}, \hat{p}^{(j)}$, and $\hat{q}^{(j)}$, the EM algorithm is as follows:

E-Step:

$$E[\log L_c|x;\hat{\mu}_1^{(j)},\hat{\mu}_2^{(j)},\hat{p}^{(j)},\hat{q}^{(j)}] = [\log p + \log f(x|n;\mu_1,\rho)]\hat{\omega}^{(j)}(n) + \sum_{s=1}^{n-1}[\log(1-p) + \log q(s) + \log f(x|s;\mu_1,\mu_2,\rho)]\hat{\omega}^{(j)}(s)$$

M-Step:

The initial value for $q(s) = 1/(n-1),\ s = 1,2,\ldots,n-1$. Maximizing the above equation with respect to $q(s),\ s = 1,2,\ldots,n-1$, subject to $\sum_{s=1}^{n-1} q(s) = 1$, gives

$$\hat{q}^{(j+1)}(t) = \frac{\hat{\omega}^{(j)}(t)}{\sum_{s=1}^{n-1}\hat{\omega}^{(j)}(s)} = \frac{\hat{q}^{(j)}(t)\hat{f}^{(j)}(x|t)}{\sum_{s=1}^{n-1}\hat{q}^{(j)}(s)\hat{f}^{(j)}(x|s)}.$$

For the exponential family with density $f(x;\mu) = h(x)\exp\{T(x)\mu - \kappa(\mu)\}$ the procedure for maximization using the EM algorithm is as follows:

$$\log f(x|t;\mu_1,\mu_2) = \sum_{i=1}^{n} h(x_i) + \sum_{i=1}^{t} T(x_i)\mu_1 + \sum_{i=t+1}^{n} T(x_i)\mu_2 - t\kappa(\mu_1) - (n-t)\kappa(\mu_2)$$

The updated estimates of $\hat{\mu}_1^{(j+1)}$ and $\hat{\mu}_2^{(j+1)}$ are the solutions of

$$\kappa'(\mu_1) = \frac{\sum_{t=1}^{n} S_t\hat{\omega}^{(j)}(t)}{\sum_{t=1}^{n} t\hat{\omega}^{(j)}(t)} = \frac{(1-p)\sum_{t=1}^{n-1}\hat{q}^{(j)}(t)S_t\hat{f}^{(j)}(x|t) + pS_n\hat{f}^{(j)}(x|n)}{(1-p)\sum_{t=1}^{n-1}\hat{q}^{(j)}(t)t\hat{f}^{(j)}(x|t) + pn\hat{f}^{(j)}(x|n)}$$

and

$$\kappa'(\mu_2) = \frac{\sum_{t=1}^{n-1}(S_n - S_t)\hat{\omega}^{(j)}(t)}{\sum_{t=1}^{n}(n-t)\hat{\omega}^{(j)}(t)} = \frac{\sum_{t=1}^{n-1}\hat{q}^{(j)}(t)(S_n - S_t)\hat{f}^{(j)}(x|t)}{\sum_{t=1}^{n-1}\hat{q}^{(j)}(t)(n-t)\hat{f}^{(j)}(x|t)}$$

In the case of a nonparametric assumption for $\omega(\tau)$, we assume $q(s)$ has a uniform prior and μ_1 is known. In the case of a parametric assumption for $\omega(\tau)$, we assume a binomial prior for $q(s)$, $s = 1, 2, \ldots, n-1$, and μ_1 can be estimated if it is unknown. The binomial prior for $q(s)$ can be defined as

$$q(s;\gamma) = \binom{n-2}{s-1}\gamma^{s-1}(1-\gamma)^{n-s-1}, \quad s = 1, 2, \ldots, n-1 \tag{4.4}$$

where γ $(0 < \gamma < 1)$ is unknown. The iterative solution for the maximum likelihood estimator of γ is given by:

$$\hat{\gamma}^{(j+1)} = \frac{\sum_{s=1}^{n-1}(s-1)\hat{\omega}^{(j)}(s)}{(n-2)\sum_{s=1}^{n-1}\hat{\omega}^{(j)}(s)} = \frac{\sum_{s=1}^{n-1}(s-1)\binom{n-2}{s-1}(\hat{\gamma}^{(j)})^{s-1}(1-\hat{\gamma}^{(j)})^{n-s-1}\hat{f}^{(j)}(x|s)}{(n-2)\sum_{s=1}^{n-1}\binom{n-2}{s-1}(\hat{\gamma}^{(j)})^{s-1}(1-\hat{\gamma}^{(j)})^{n-s-1}\hat{f}^{(j)}(x|s)} \tag{4.5}$$

and the estimates of μ_1 and μ_2 are obtained through the procedure discussed earlier except for $\hat{q}^{(j)}(s)$, which has to be replaced by $q(s;\hat{\gamma}^{(j)})$.

If we are considering normally distributed data, then the updated estimates $\hat{\mu}_1^{(j+1)}, \hat{\mu}_2^{(j+1)}$ are as follows:

$$\hat{\mu}_1^{(j+1)} = \frac{(1-p)\sum_{t=1}^{n-1} S_t \hat{q}^{(j)}(t) \hat{f}^{(j)}(x|t) + pS_n \hat{f}^{(j)}(x|n)}{(1-p)\sum_{t=1}^{n-1} t\hat{q}^{(j)}(t) \hat{f}^{(j)}(x|t) + pn\hat{f}^{(j)}(x|n)}$$

$$\hat{\mu}_2^{(j+1)} = \frac{\sum_{t=1}^{n-1} (S_n - S_t)\hat{q}^{(j)}(t) \hat{f}^{(j)}(x|t)}{\sum_{t=1}^{n-1} (n-t)\hat{q}^{(j)}(t) \hat{f}^{(j)}(x|t)}$$

where $S_t = \sum_{s=1}^{t} x_s$ and

$$\hat{f}^{(j)}(x|t) = (2\pi\hat{\sigma^2}^{(j)})^{-n/2} exp\{-\left[\sum_{s=1}^{t}(x_s - \hat{\mu}_1^{(j)})^2 + \sum_{s=t+1}^{n}(x_s - \hat{\mu}_2^{(j)})^2\right]/2\hat{\sigma^2}^{(j)}\}.$$

The iterative estimate of $\hat{\sigma^2}^{(j+1)}$ is given by

$$\hat{\sigma^2}^{(j+1)} = \frac{1}{n}\sum_{s=1}^{n} \hat{\omega}^{(j)}(s)\left[\sum_{j=1}^{s}(x_j - \hat{\mu}_1^{(j+1)})^2 + \sum_{j=s+1}^{n}(x_j - \hat{\mu}_2^{(j+1)})^2\right]$$

This Bayesian approach is computationally expensive since it involves the EM procedure. In general, the EM procedure is quite slow and heavily depends on the initial approximate values of the parameters. In the case of a nonparametric prior for $q(s)$, Bansal et al. (2008) suggested initial values of $q(s) = 1/(n-1), s = 1, 2, \ldots, n-1; \mu_1 = \mu_2 = 0$; $\sigma^2 = 1$; and $p = 0.05$ (for the normally distributed change point model). We are interested in using the EM test for homogeneity suggested by Li et al. (2009) and Chen and Li (2009) to accelerate the Bayesian approach. We modified

the initial value of $q(s)$ in the Bayesian approach to accelerate the identification of the change point.

4.2 EM Test

To accelerate and ensure convergence of the EM algorithm, Li et al. (2009) and Chen and Li (2009) proposed the EM test for homogeneity in finite mixture models. Let $X_1, X_2, \ldots, X_n$ be a random sample from a model $f(x;\mu)$. Its log-likelihood function is given by

$$\ell_n(\beta, \mu_1, \mu_2) = \sum_{i=1}^{n} \log\{(1-\beta)f(x_i;\mu_1) + \beta f(x_i;\mu_2)\} \tag{4.6}$$

where β $(0 \leq \beta \leq 1)$ and $1-\beta$ are the mixing proportions, and μ_1 and μ_2 are the mixing parameters. Then the penalized likelihood function is defined to be

$$PL_n(\beta, \mu_1, \mu_2) = \ell_n(\beta, \mu_1, \mu_2) + p(\beta)$$

where $p(\beta)$ is a penalty function of β. This test procedure is based on the EM-algorithm. For a fixed $\beta = \beta_0 \in (0, 0.5]$, the penalized likelihood ratio statistic is defined to be

$$M_n(\beta_0) = 2\{PL_n(\beta_0, \tilde{\mu}_{01}, \tilde{\mu}_{02}) - PL_n(0.5, \tilde{\mu}_0, \tilde{\mu}_0)\} \tag{4.7}$$

where $PL_n(\beta_0, \mu_1, \mu_2)$ is maximized at $\tilde{\mu}_{01}$ and $\tilde{\mu}_{02}$ and $PL_n(0.5, \mu_1, \mu_2)$ has a maximum at $\tilde{\mu}_0$. $M_n(\beta_0)$ has a χ^2-type limiting distribution even if assumptions 1 and 2 (defined in Li et al. (2009)) are not satisfied. If the data follow a different model, β is not the same as β_0 and the test based on (4.7) can become inefficient. The authors solved this problem by updating the values of β using the EM-algorithm. Using the EM-algorithm, they update both the mixing proportion β and the mixing parameters (μ_1, μ_2). They proposed the EM test to achieve better efficiency than that of $M_n(\beta_0)$ by updating the value of β_0. They used a number of initial values for β_0 to speed up the process and reduced the number of iterations (m) to capture the true value of μ even if the data follow a different model. They define the test statistic $EM_n^{(m)} = \max_{1 \le j \le J} \{M_n^{(m)}(\beta_{0j})\}$, where J is the number of initial values of β_0 and m is the fixed number of iterations. They prove, under general conditions, that for a fixed m and a pre-chosen β_{0j} the test statistic $EM_n^{(m)}$ has a limiting distribution $0.5\chi_0^2 + 0.5\chi_1^2$. A detailed discussion of the EM test is given in Li et al. (2009) and Chen and Li (2009).

We propose using the EM test in the Bayesian approach to the change point problem suggested by Bansal et al. (2008). For the EM test procedure, we divide the sample into different groups (K) and set $q(s)$ high for one group and low for all the remaining groups and perform the EM algorithm for a fixed number of iterations

(m). The data point with the highest $q(s)$ value is identified as the candidate change point and corresponding likelihood value is also computed. We repeat this K times. Each time, one group has a high $q(s)$ value and the others have low $q(s)$ values. From these K candidate change points, the one with maximum likelihood value is identified as the change point. The basic idea is that if the true change point is in the group with a large initial value for $q(s)$, within a few iterations it will have a $q(s)$ value close to one and correctly identify the change point.

4.2.1 Step-by-Step EM Test Procedure for Change Point Model

1. Divide the data into K subgroups (i.e., number of EM tests) and set the number of iterations for each EM test to m (we considered 5 and 10).

2. Set the $q(s)$ values high for all the data points in one subgroup and low for all points in the remaining subgroups such that the $q(s)$ sum is one $(s = 1, 2, \ldots, n-1)$.

3. Perform the EM procedure for m iterations and select the data point with maximum $q(s)$ as a candidate for the change point.

4. For the candidate change point, compute the likelihood value.

5. Repeat steps 2–4 K times, and each time assign the high $q(s)$ value to a different subgroup.

6. From the K candidate change points, identify the data point with the maximum likelihood value as the change point.

4.3 Simulation Studies for Normally Distributed Data (Univariate)

4.3.1 Case 1: Data With No Change Point for Mean

In the case of no change point (the null case), following Remark 2.1 of Bansal et al. (2008) we checked the value of $\hat{\omega}^{(j)}(n)$ using (4.3) at the final iteration. If the value is greater than an appropriate constant d, we conclude that there is no change point. No specific value for d is provided by Bansal et al. (2008). We simulated 10000 samples of sizes $n = 30$ and $n = 10$ with no change point and estimated the $\hat{\omega}^{(j)}(t)$ values. We found that $\hat{\omega}^{(j)}(n)$ is the second or third largest. Using EM test procedures we found that $\hat{\omega}^{(j)}(n)$ is within the top 30% of all $\hat{\omega}^{(j)}(t)$ values after m iterations ($m = 5$ or 10).

4.3.2 Case 2: Data With Change Point for Mean

To simplify the comparison, we adopted the simulation approach used in Bansal et al. (2008). For normally distributed data of size $n = 10$, we consider three change points (τ) at 3, 5, and 8 respectively with some shift in the mean, say 0.5. We simulate the first τ observations from $N(0, 1)$ and the remaining $(n - \tau)$ from $N(0.5, 1)$. In the EM test we used $m = 5$ iterations and $K = 2$ subgroups. We set $q(s)$ for one subgroup to 0.8 and the $q(s)$ sum for the other group to 0.2, i.e., we divided the data set into two subgroups of size five. Thus, for the first EM-test, $\sum_{s=1}^{5} q(s) = 0.8$ and $\sum_{s=6}^{9} q(s) = 0.2$. Note that for a sample of size n, we need not consider $q(n)$. For the second EM-test, $\sum_{s=1}^{5} q(s) = 0.2$ and $\sum_{s=6}^{9} q(s) = 0.8$. Within each subgroup, we assigned the $q(s)$ values equally. For each EM test, a candidate change point is identified based on the maximum of the $q(s)$ values. For each candidate change point, we compute the likelihood as if it is the change point, and the candidate change point with the maximum likelihood value (among the K EM tests) is identified as the change point. We performed 10000 simulations and plotted the proportion of times each data point is identified as the change point. For comparison, we also employed the method proposed in Bansal et al. (2008). We also considered larger shifts in the mean (1 and 1.5). These results are given in Figs. 4.1, 4.2, and 4.3 for $\tau = 3, 5$, and 8

respectively. In all cases, we found that even for $m = 5$, the EM test procedure gives similar results to the full EM algorithm (Bansal et al. (2008)) for identifying the change point. From these figures, it is clear that the proportion of times the true change point is identified as the change point by both methods are almost the same. Also note that the proportion of times each data point is identified as the change point also has similar pattern by both methods. We also used $m = 10$ iterations for the EM-test procedure. Since the results are similar for $m = 5$ and $m = 10$, we omit the results for $m = 10$ here.

We also consider a larger sample size, $n = 30$, with change points (τ) at 6, 15, and 24. Here, we set the number of subgroups $K = 6$. We used the same simulation set-up of $n = 10$ for the data generation. The results are depicted in Figs. 4.4, 4.5, and 4.6, and these graphs clearly show the similar performance of the Bayesian approach with the EM test and the Bayesian approach with the full EM algorithm. Our simulation study clearly indicates that the EM test performs as well as the full Bayesian approach for change point problems, and it requires fewer iterations.

4.3.3 Case 3: Binomial Prior for $q(s)$

In the parametric prior case, the EM test of the Bayesian approach and the EM test for homogeneity are similar. Instead of using a uniform prior for q, we

suggest a binomial prior in (4.4), and the iterative solution for the maximum likelihood estimator of γ is given by (4.5). The estimates of θ_1 and θ_2 are obtained through the procedure discussed earlier except for $\hat{q}^{(m)}(s)$, which is replaced by $q(s;\hat{\gamma}^{(m)})$. In the simulation study with the EM test we used a binomial prior with $\gamma = 0.2, 0.5$, and 0.8 and, for each prior, identified the candidate change point as $n\hat{\gamma}^{(m)}$ where $\hat{\gamma}^{(m)}$ is the estimate of γ in the final iteration. We computed the likelihood corresponding to each candidate change point with $q(n\hat{\gamma}^{(m)}) = 1$. The shift in the mean is set to 1. Using the EM test, we estimate the change point for each prior and the point with the maximum likelihood is identified as the change point. For the simulation, we consider a sample size of $n = 10$ with change points at 3, 5, and 8. Bansal et al. (2008) did not fully develop a procedure for the identification of the change point. Therefore, we compared the results for full iteration of the EM test and the EM test with $m = 5$. We computed the proportion of times each point was identified as the change point based on 10000 simulations. The results are given in Fig. 4.7. From Fig. 4.7, we see that the EM test performs as well as the EM with full iteration. From this figure, it is clear that EM test with $m = 5$ has slightly high proportion of times the true change point is identified as compared to the EM test with full iteration. Also note that the proportion of times each data point is identified as the change point is also has similar pattern by both methods.

We also conducted simulation studies for a larger sample size, $n = 30$, with change points at 6, 15, and 24. We simulated data from a normal distribution with a unit shift in mean at the change points. A summary of the results is given in Fig. 4.8. We see from Fig. 4.8 that the EM test performs as well as the EM with full iteration.

4.4 Performance of MIC, ELIC, and Bayesian approaches

We compared the performance of the Bayesian approach for the change point problem with MIC and ELIC. We restricted our comparison to the nonparametric-prior case in both the Bayesian approach and the EM tests. Since there is no well-defined procedure in the Bayesian approach to signal the change point, we identified the position where the change point was detected and compared it with that of MIC and ELIC. For MIC and ELIC, we identified the data point corresponding to the test statistics S_n and $ELIC(\tau)$ and used that for the comparison. We compute the mean and standard deviation of the identified change points for the comparison. We considered sample size $n = 200$ and two locations of the change points as per the performance studies carried out in Chapter 3. In the first scenario, we generated the first 100 observations from $N(0, 1)$ and the remaining 100 observations from $N(\mu, 1)$.

We assumed shifts in the mean of μ=0.2, 0.4, 0.6, 0.8, and 1, and set the number of subgroups to $K = 5$. Summary statistics based on 5000 simulations are given in Table 4.1. We see from Table 4.1 that, on average, all the methods correctly identified the change point. Note that the standard deviation for ELIC and MIC is slightly lower than that for the Bayesian approaches.

Method	Quantity	Shift in Mean				
		0.2	0.4	0.6	0.8	1
Bayesian	Mean	102.52	101.95	101.02	100.71	100.41
	SD	40.93	24.09	13.38	8.29	5.39
Bayesian EM test, m=5	Mean	101.4	101.24	100.24	100.05	99.97
	SD	51.25	35.58	20.39	10.59	6.15
MIC	Mean	100.17	100.28	100.14	100.08	99.99
	SD	35.44	24.02	14.37	8.9	5.69
ELIC	Mean	99.95	100.12	100.13	100.18	99.94
	SD	33.54	22.63	13.95	8.9	6.12

Table 4.1: Mean and standard deviation of location of change point for normal data with $\tau = 100$

For the second scenario, with the change point at 150, summary statistics based

on 5000 simulations are given in Table 4.2. Here also all the methods identified the change point correctly. Note that the standard deviation for ELIC is low compared to those of the other methods. We see from Tables 4.1 and 4.2 that the standard deviation of location of change points for all methods is less when the change point is close to the middle of the data.

Method	Quantity	Shift in Mean				
		0.2	0.4	0.6	0.8	1
Bayesian	Mean	116.8	136.49	144.77	148.04	149.15
	SD	46.96	32.94	19.56	10.97	6.75
Bayesian EM test, m=5	Mean	111.66	133.32	144.78	148.46	149.55
	SD	55.84	41.95	24.63	13.32	6.68
MIC	Mean	111.84	131.76	143.04	147.46	149.06
	SD	41.98	34.1	21.37	11.99	6.53
ELIC	Mean	110.81	130.16	140.43	144.32	145.5
	SD	40.11	32.64	21.78	13.26	9.13

Table 4.2: Mean and standard deviation of location of change point for normal data with $\tau = 150$

4.4.1 Non-normal data

We consider data from an exponential distribution and assume a normal distribution, as many practitioners do. Thus, we have misspecified exponential data with a normal set-up for the location of the change point. We consider the two scenarios discussed in the previous section. In the first scenario, the first 100 observations are generated from exp(1), where exp(1) is an exponential distribution with mean 1, and the remaining 100 observations from exp(μ), i.e., data from an exponential distribution with mean μ. We considered $\mu = 0.9, 0.8, 0.7, 0.6$, and 0.5, i.e., a downward shift in the mean of 0.1, 0.2, 0.3, 0.4, and 0.5. We simulated 5000 data sets and computed the mean and standard deviation of the identified change points for all the methods. A summary of the results is given in Table 4.3. From Table 4.3, we see a significant difference between ELIC and the other methods. For ELIC, the mean change point is steady near 100 for all cases irrespective of the shift in the mean. For the other methods, the change point is located well below 100.

In the second scenario with the change point at 150, we simulated 5000 data sets with sample size $n = 200$ and the same downward shift. We computed the mean and standard deviation of the location of the change points. A summary of the results is given in Table 4.4. The EM test is better at locating the change point, but the standard deviation is high. ELIC performs better than the other methods except the

Method	Quantity	Shift in Mean				
		0.1	0.2	0.3	0.4	0.5
Bayesian	Mean	98.81	95.04	94.27	94.3	95.08
	SD	47.39	38.46	27.44	17.89	11.94
Bayesian EM test, m=5	Mean	132.38	113.07	96.89	90.43	91.32
	SD	68.53	62.53	46.87	29.69	18.93
MIC	Mean	95.28	91.51	90.18	90.83	92.72
	SD	41.23	35.32	28.69	21.9	16.08
ELIC	Mean	101.28	101.43	101.76	101.43	100.94
	SD	42.57	36.55	28.36	20.72	14.02

Table 4.3: Mean and standard deviation of location of change point for exponential data with $\tau = 100$

EM test, and the standard deviation is small.

Method	Quantity	Shift in Mean				
		0.1	0.2	0.3	0.4	0.5
Bayesian	Mean	103.7	112.91	123.56	133.09	138.78
	SD	48.95	45.48	38.51	29.26	21.04
Bayesian EM test, m=5	Mean	138.22	131.61	129.6	130.03	135.72
	SD	66.96	62.63	53	42.74	31.99
MIC	Mean	99.59	105.73	116	126.15	134.42
	SD	41.65	41.03	39.26	34.57	27.06
ELIC	Mean	105.7	117.16	130.02	140.15	145.26
	SD	44.8	43.44	36.83	26.26	15.12

Table 4.4: Mean and standard deviation of location of change point for exponential data with $\tau = 150$

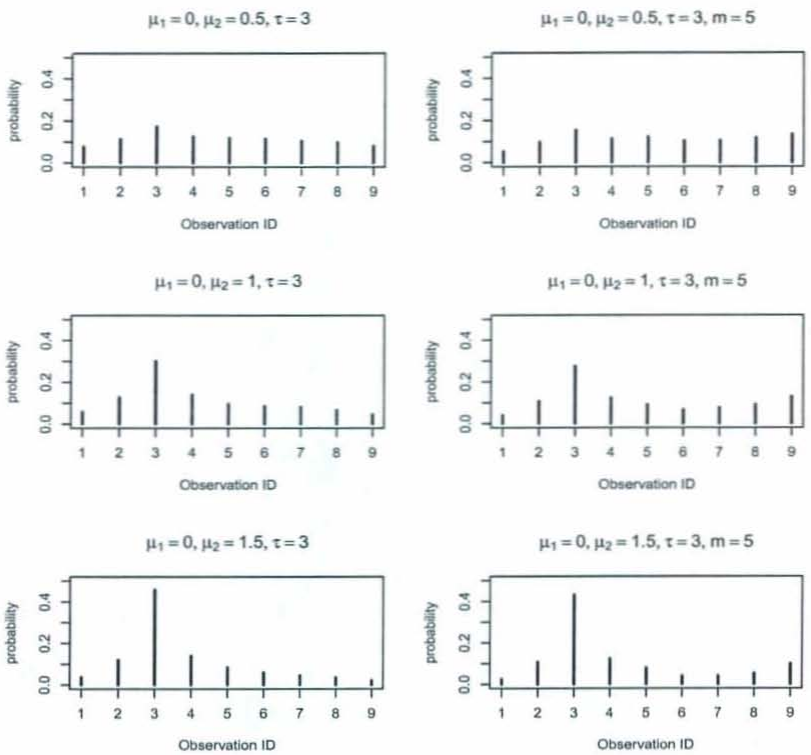

Figure 4.1: Signal probability for each data point as a change point for sample size 10. First figure in each panel refers to Bansal et al. procedure; second figure refers to EM test based on 5 iterations. First panel is for shift 0.5, second panel for shift 1.0, and third panel for shift 1.5 with change point at 3rd position.

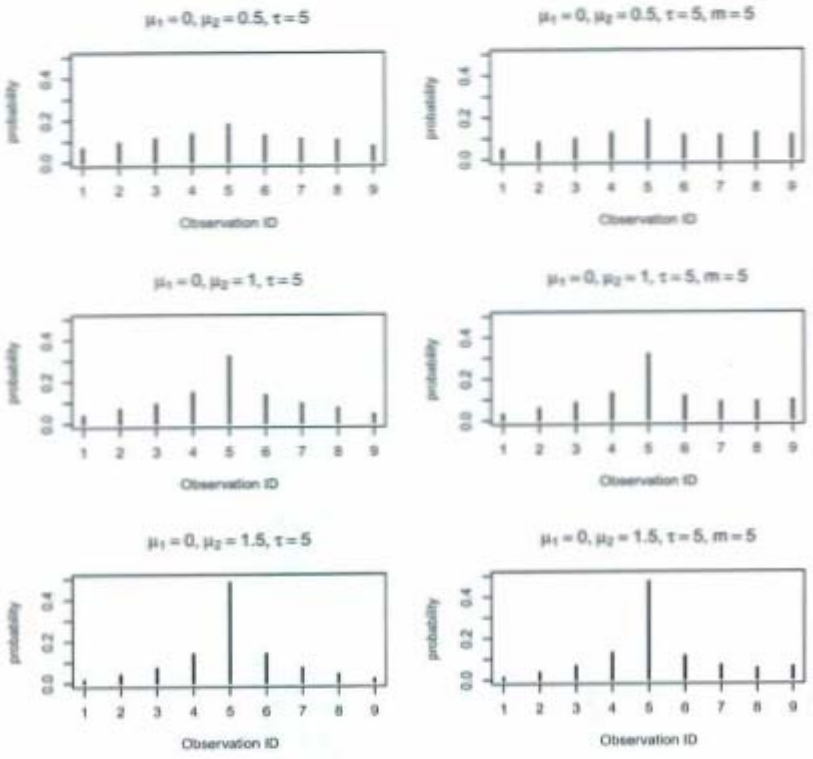

Figure 4.2: Signal probability for each data point as a change point for sample size 10. First figure in each panel refers to Bansal et al. procedure; second figure refers to EM test based on 5 iterations. First panel is for shift 0.5, second panel for shift 1.0, and third panel for shift 1.5 with change point at 5th position.

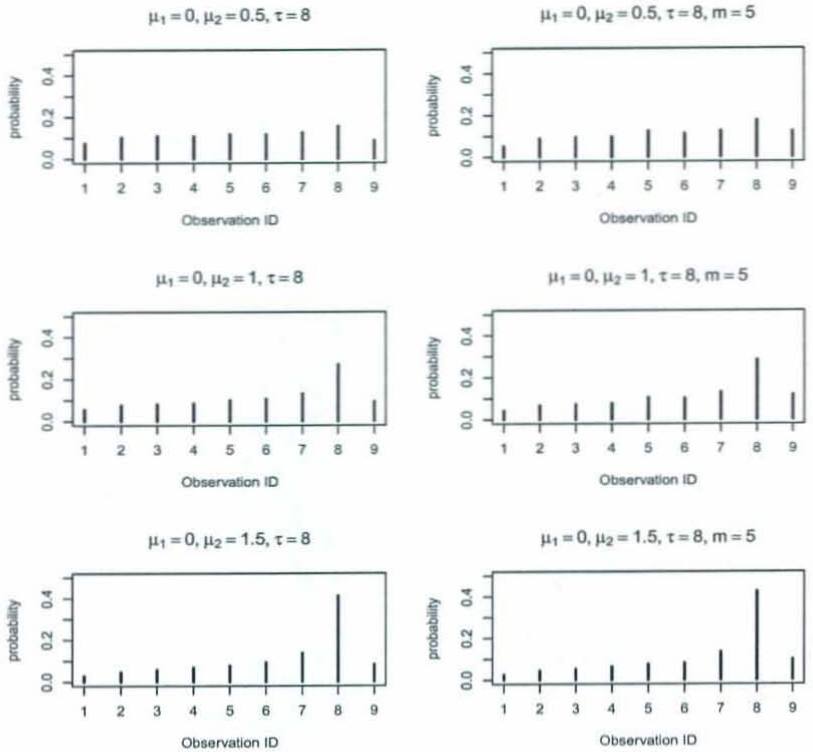

Figure 4.3: Signal probability for each data point as a change point for sample size 10. First figure in each panel refers to Bansal et al. procedure; second figure refers to EM test based on 5 iterations. First panel is for shift 0.5, second panel for shift 1.0, and third panel for shift 1.5 with change point at 8th position.

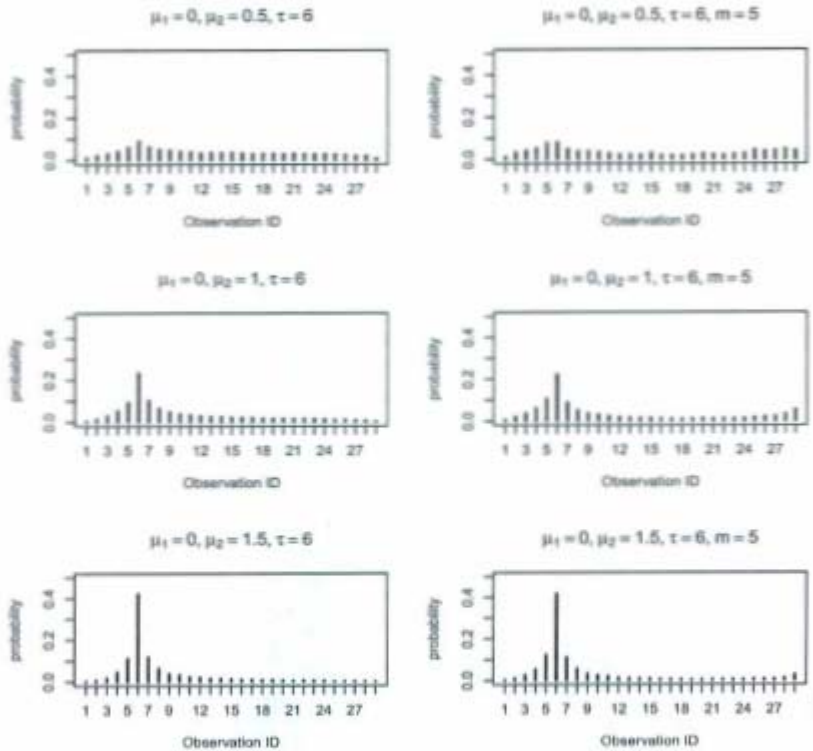

Figure 4.4: Signal probability for each data point as a change point for sample size 30. First figure in each panel refers to Bansal et al. procedure; second figure refers to EM test based on 5 iterations. First panel is for shift 0.5, second panel for shift 1.0, and third panel for shift 1.5 with change point at 6th position.

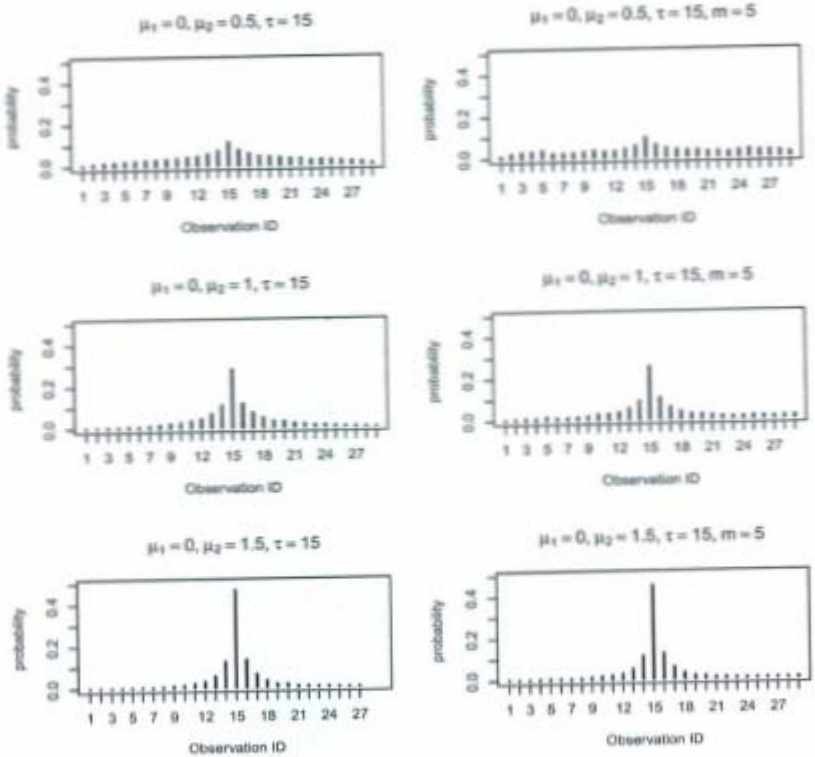

Figure 4.5: Signal probability for each data point as a change point for sample size 30. First figure in each panel refers to Bansal et al. procedure; second figure refers to EM test based on 5 iterations. First panel is for shift 0.5, second panel for shift 1.0, and third panel for shift 1.5 with change point at 15th position.

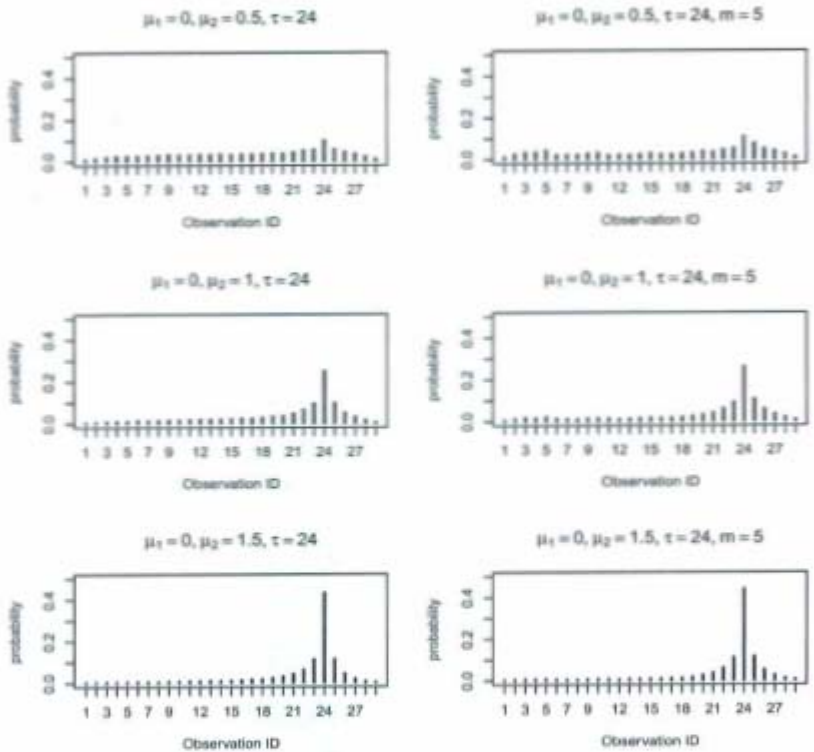

Figure 4.6: Signal probability for each data point as a change point for sample size 30. First figure in each panel refers to Bansal et al. procedure; second figure refers to EM test based on 5 iterations. First panel is for shift 0.5, second panel for shift 1.0, and third panel for shift 1.5 with change point at 24th position.

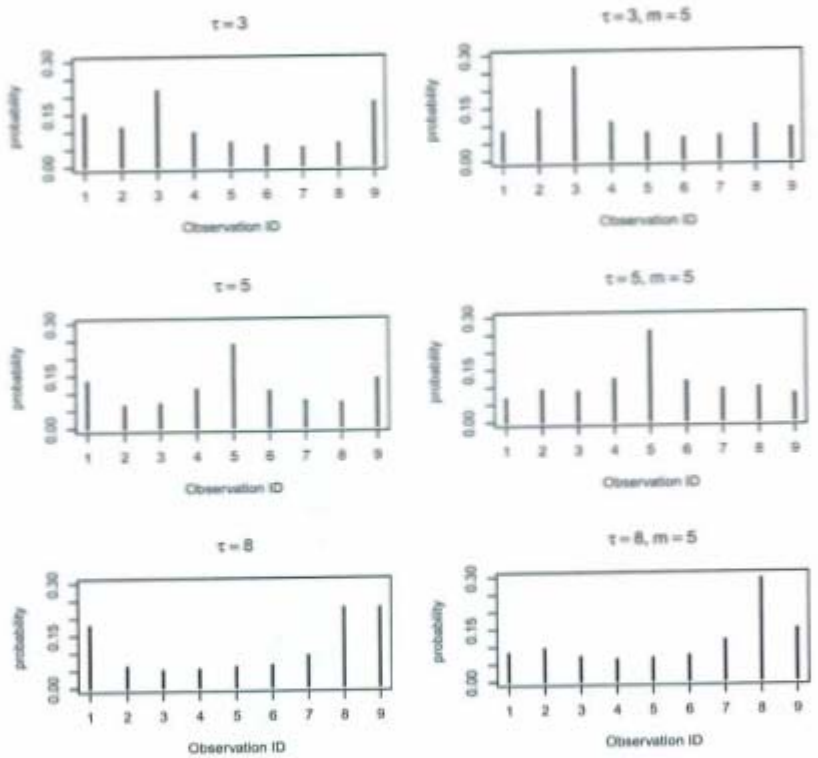

Figure 4.7: Signal probability for each data point as a change point for sample size 10. First figure in each panel refers to full iteration with EM test; second figure refers to EM test based on 5 iterations. First panel is for change point at 3, second figure for change point at 5, and third figure for change point at 8. Shift is fixed to 1.

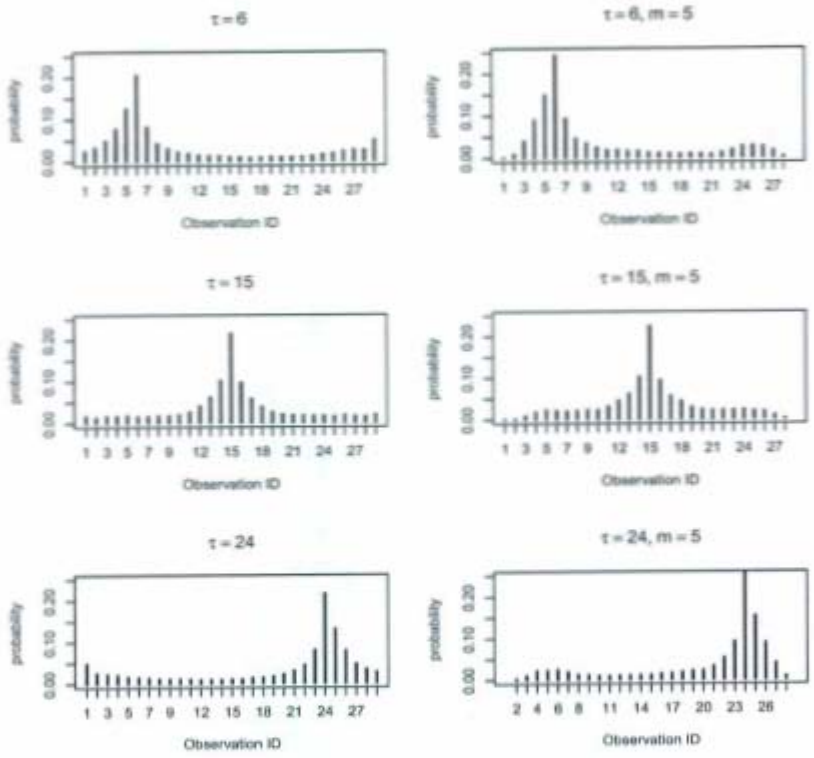

Figure 4.8: Signal probability for each data point as a change point for sample size 30. First figure in each panel refers to full iteration with EM test; second figure refers to EM test based on 5 iterations. First panel is for change point at 6, second figure for change point at 15, and third figure for change point at 24. Shift is fixed to 1.

Chapter 5

Case Studies

In this chapter we apply the methods developed in this thesis to two data sets. We do not know the true change points for these data sets, but our analysis will help us to understand how our methods perform at identifying the change points.

5.1 Example 1: Runout data

In this section, we discuss the identification of the change point for the circular runout (ovality) data set discussed in Chapter 1. The quality characteristic of interest is the runout on the head-stock side of an engine valve. The lower and upper limits on the runout, set by the customer, are zero microns and 0.5 microns respectively.

If the runout is more than 0.5 microns, the product must be reworked or rejected. Therefore, runout must be monitored on a continuous basis. As discussed earlier, a histogram of the historical data of ovality (see Fig. 5.1) indicates that its distribution is non-normal (slightly right-skewed). An empirical likelihood based on a control chart is proposed to monitor this process (see Variyath (2010)). Since the small changes in the process mean must be monitored cautiously, a change point analysis is recommended. Since it is not easy to approximate the distribution of the ovality, ELIC is the most appropriate method. The ovality data for 100 consecutive shafts are plotted in Fig. 5.2.

We applied all five change-point methods, the change point model, MIC, ELIC, the Bayesian approach, and the Bayesian approach with the EM test ($m = 5$). All the methods conclude that there is a shift at the 61^{st} position.

5.2 Example 2: Chemical data

We identify the change point for the chemical process concentration data reported in Box, Jenkins and Reinsel (1994). A plot of this data set (number of observations=197) is given in Fig. 5.4, and a histogram is given in Fig. 5.3. The histogram clearly indicates that the distribution is non-normal (slightly right-skewed). From

Fig. 5.4, we suspect that there are multiple change points, around 100 and 150. Since it is not easy to approximate the distribution of the concentration readings, ELIC may be the most appropriate method to identify the change point.

We applied all five methods (Bayesian, Bayesian with EM test and m=5, Hawkins method, MIC, ELIC) to detect change points. All the methods except ELIC conclude that there is a shift at the 170^{th} position. Only ELIC identified the early shift at the 150^{th} position. When we checked the test statistics for all the data points, we realized there is also a shift at data point 55. However, the maximum values of the test statistics occur only at the 150^{th} and 170^{th} positions, and only ELIC detected the shift at data point 150.

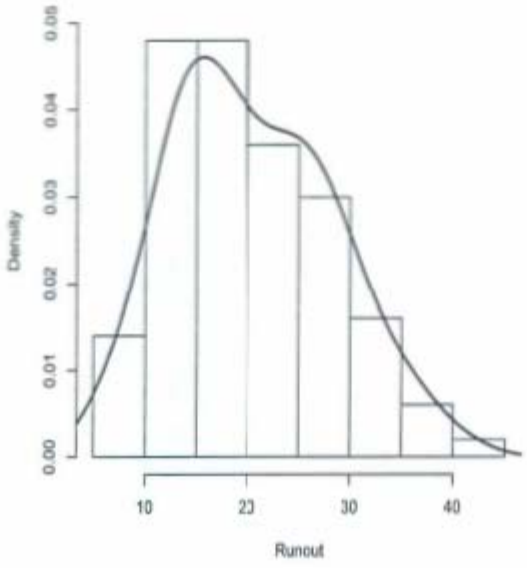

Figure 5.1: Histogram and density plot of runout data

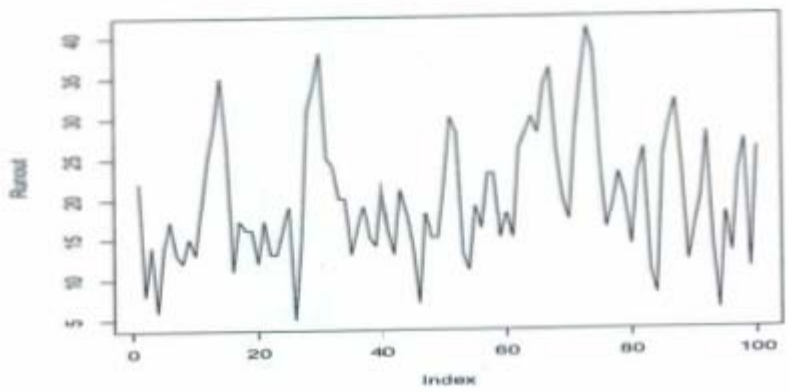

Figure 5.2: Plot of runout data

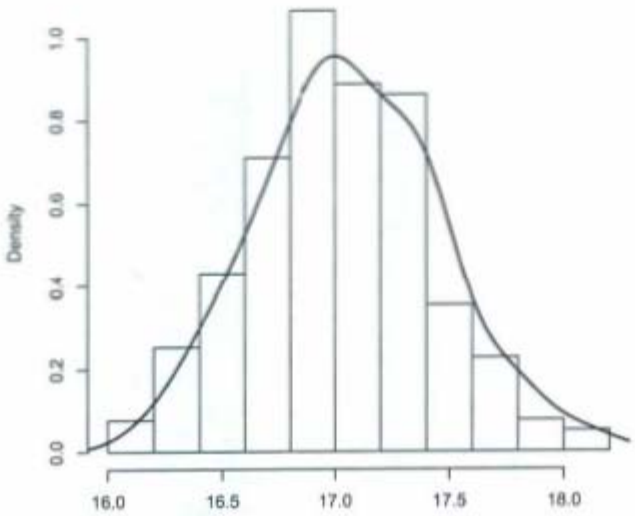

Figure 5.3: Histogram and density plot of chemical data

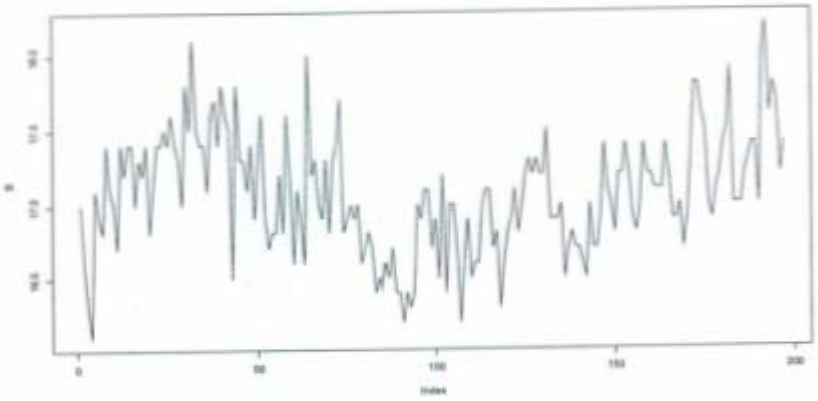

Figure 5.4: Plot of chemical data

Chapter 6

Concluding Remarks

Quality has become an important consumer decision factor. All modern quality-management practices aim to achieve quality with minimum variation. A shift in the process can diminish the quality. To detect a shift in the process mean (the change point), practitioners usually prefer conventional methods such as Shewart control charts, CUSUM charts, or EWMA charts. Shewart control charts are not able to detect small shifts, and CUSUM and EWMA charts are also slow to detect larger shifts. Better methods are needed for change point analysis. Hawkins et al. (2003) proposed a change point analysis for the normal model based on the two-sample Student's t statistic. Chen et al. (2006) suggested a modified information criterion (MIC) based on the likelihood ratio test and Bayesian information criterion (BIC).

However, these methods rely on parametric assumptions for the quality characteristics. If we are unsure about the distributional properties of the quality characteristic, we usually prefer a normal model assumption. However, it can lead to an incorrect conclusion. We therefore need to develop a nonparametric procedure for change point analysis. We propose an EL-based information criterion (ELIC) to identify changes in the process mean. The main advantage of ELIC is that we do not need to specify the distribution of the quality characteristic. Using the adjusted empirical likelihood, we developed a computational algorithm to compute ELIC. Our simulation studies clearly indicate that ELIC performs as well as the parametric-based methods, if the distribution of the quality characteristic is known. When the distribution is misspecified or approximated, parametric methods failed to detect the change detected by ELIC.

In the context of change point analysis, Bansal et al. (2008) proposed a semi-Bayesian approach with the EM algorithm. This method is computationally expensive because of the use of the EM algorithm. To accelerate this Bayesian approach, we suggest using the EM test proposed by Li et al. (2009) and Chen and Li (2009). We used the EM test with a specific initial value of the prior probability q to accelerate the identification of the change point. Under a binomial-prior assumption for q, we modified the Bayesian approach by using different binomial-prior assumptions. In

both cases, we used the EM test with a fixed number of iterations, $m = 5$. Simulation studies show that the EM test of the Bayesian approach works as well as the Bayesian approach with full iteration.

We compared the performance of the Bayesian approach, the Bayesian approach with the EM test, MIC, and ELIC. The results show that ELIC performs well compared to other methods in both specified and misspecified scenarios for the distributional assumption of the quality characteristic. We applied our proposed methods to real case studies, the circular runout (ovality) data set and the chemical process concentration data available in Box et al. (1994).

In this thesis, we have developed the theory for multivariate process mean. But we presented the simulation study for univariate process mean only. We would like to extend ELIC to the situations when there is a change in mean or variance or both. We would like to extend our simulation studies to identify the changes in multivariate quality characteristic as well. In this thesis, we considered the situation where there is only one change point. But in practise, we may come across situations where there are multiple change points (see chemical data set discussed in Chapter 5). So we would like to extend our work to detect multiple change points as well.

Bibliography

[1] Bansal, N.K., Du, H., and Hamedani, G.G. (2008). An Application of EM Algorithm to a Change-Point Problem. *Communications in Statistics: Theory and Methods*, **37**, 2010-2021.

[2] Box, E.P.G., Jenkins, G.M., and Reinsel, G.C. (1994). Time Series Analysis Forecasting and Control. *Pearson Education.*

[3] Chen, J., Gupta, A.K., and Pan, J. (2006). Information Criterion and Change Point Problem for Regular Models. *Sankhya: The Indian Journal of Statistics*, **68**, 252-282.

[4] Chen, J. and Li, P. (2009). Hypothesis Test for Normal Mixture Models: The EM approach. *The Annals of Statistics*, **37**, 2523-2542.

[5] Chen, J., Sitter, R.R., and Wu, C. (2002). Using Empirical Likelihood Methods to Obtain Range Restricted Weights in Regression Estimators for Surveys.

Biometrika, **89**, 230-237.

[6] Chen, J., Variyath, A.M., and Abraham, B. (2008). Adjusted Empirical Likelihood and its Properties. *Journal of Computational and Graphical Statistics*, **17**, 426-443.

[7] Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum Likelihood from Incomplete Data via EM Algorithm (with Discussion). *Journal of Royal Statistical Society*, **B 39**, 1-38.

[8] Hawkins, D.M. (1977). Testing a Sequence of Observations for a Shift in Location. *Journal of the American Statistical Association*, **72**, 180-186.

[9] Hawkins, D.M. and Olwell, D.H. (1998). Cumulative Sum Charts and Charting for Quality Improvement. *Springer-Verlag New York, Inc.*

[10] Hawkins, D.M., Qiu, P., and Kang, C.W. (2003). The Changepoint Model for Statistical Process Control. *Journal of Quality Technology*, **35**, 355-366.

[11] Hawkins, D.M. and Zamba, K.D. (2005). Statistical Process Control for Shifts in Mean or Variance Using a Changepoint Formulation. *Technometrics*, **47**, 164-173.

[12] Jing, B.Y. (1995). Two-Sample Empirical Likelihood Method. *Statistics & Probability Letters*, **24**, 315-319.

[13] Li, P., Chen, J., and Marriott, P. (2009). Non-Finite Fisher Information and Homogeneity: An EM Approach. *Biometrika*, **96**, 411-426.

[14] Liu, Y., Zou, C., and Zhang, R. (2008). Empirical Likelihood for the Two-Sample Mean Problem. *Statistics & Probability Letters*, **78**, 548-556.

[15] NIST/SEMATECH e-Handbook of Statistical Methods. Retrieved from http://www.itl.nist.gov/div898/handbook/

[16] Owen, A.B. (1988). Empirical Likelihood Ratio Confidence Interval for a Single Functional. *Biometrika*, **75**, 237-249.

[17] Owen, A.B. (2001). Empirical Likelihood. *Chapman & Hall/CRC.*

[18] Page, E.S. (1954). Continuous Inspection Schemes. *Biometrika*, **41**, 100-115.

[19] Roberts, S.W. (1959). Control Chart Tests Based On Geometric Moving Averages. *Technometrics*, **1**, 239-250.

[20] Ryan, T.P. (2000). Statistical Methods for Quality Improvement. *John Wiley & Sons, Inc.*

[21] Sen, A. and Srivastava, M.S. (1975). On Tests for Detecting Change in Mean When Variance is Unknown. *Annals of the Institute of Statistical Mathematics*, **27**, 479-486.

[22] Shewart, W.A. (1931). Economic Control of Quality of Manufactured Product. *New York: Van Nostrand.* (Reprinted in 1980 by the *American Society for Quality Control.*)

[23] Shewart, W.A. (1939). Statistical Method from the Viewpoint of Quality Control. *Washington, DC: Graduate School, Department of Agriculture* (editorial assistance by W. Edwards Deming).

[24] Variyath, A.M. (2006). Variable Selection in Generalized Linear Models by Empirical Likelihood. *Ph.D. Thesis, University of Waterloo, Canada.*

[25] Variyath, A.M. (2010). Empirical Likelihood Based Control Charts. *Submitted for publication.*

[26] Variyath, A.M., Chen, J., and Abraham, B. (2010). Empirical Likelihood Based Variable Selection. *Journal of Statistical Planning and Inference*, **140**, 971-981.

[27] Worsley, K.J. (1979). On the Likelihood Ratio Test for a Shift in Location of Normal Populations. *Journal of the American Statistical Association* **74**, 365-367.